نظریه اطلاعات، آمار و یادگیری

دکتر یاسایی

دانشگاه صنعتی شریف

مهندسی برق

برنا خدابنده ۴۰۰۱۰۹۸۹۸

تمرین شماره 2

تاریخ: 1403/02/02

$P, Q: P \ll Q \;\; X_i \sim P;\; Y_i \sim Q : Z_i = \log\left(\frac{P(X_i)}{q(X_i)}\right),\; W_i = \log\left(\frac{P(Y_i)}{q(Y_i)}\right)$ (سوال ۱

(1

$$\mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\right] \overset{\lambda \geq 0}{=} \mathbb{P}\left[e^{\lambda \sum_{i=1}^{n}(W_i - Z_i)} \geq e^{\lambda n t}\right] \leq e^{-n\lambda t} \cdot \mathbb{E}_{\substack{X_i \sim P \\ Y_i \sim Q}}\left[e^{\lambda \sum W_i - Z_i}\right]$$

$$\mathbb{E}_{\substack{X_i \sim P \\ Y_i \sim Q}}\left[e^{\lambda \sum W_i - Z_i}\right] = \mathbb{E}_{\substack{X_i \sim P \\ Y_i \sim Q}}\left[\prod_{i=1}^{n} e^{\lambda(W_i - Z_i)}\right] = \prod_{i=1}^{n} \mathbb{E}_{\substack{X_i \sim P \\ Y_i \sim Q}}\left[e^{\lambda(W_i - Z_i)}\right] = \mathbb{E}_{\substack{X_i \sim P \\ Y_i \sim Q}}\left[e^{\lambda(W_i - Z_i)}\right]^n$$

$$\Rightarrow \mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\right] \leq \exp\left(-n\lambda t + n\log \mathbb{E}\left[e^{\lambda(W_i - Z_i)}\right]\right) \qquad \log\mathbb{E}[e^{\lambda W_i}] + \log\mathbb{E}[e^{-\lambda Z_i}]$$

$$\Rightarrow \mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\right] \leq \inf_{\lambda \geq 0} \exp\Big\{-n\big(\lambda t - \underbrace{\log \mathbb{E}[e^{\lambda W_i}]}_{\psi_Q(\lambda)} - \underbrace{\log \mathbb{E}[e^{-\lambda Z_i}]}_{\psi_P(-\lambda)}\big)\Big\} = \exp(-nF(t)) \qquad \sup_{\lambda \geq 0}$$

$$\Rightarrow \boxed{\mathbb{P}\left[\sum W_i - Z_i \geq nt\right] \leq \exp(-nF(t));\; F(t) = \sup_{\lambda \geq 0}\{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)\}}$$

$$\mathbb{E}_{Y_i \sim Q} W_i = \mathbb{E}_{Y_i \sim Q} \log\left(\frac{P(Y_i)}{Q(Y_i)}\right) = -D_{KL}(Q\|P),\quad \mathbb{E}_{X_i \sim P} Z_i = \mathbb{E}_{X_i \sim P} \log\left(\frac{P(X_i)}{Q(X_i)}\right) = D_{KL}(P\|Q)$$ (2

$$F(0) = \sup_{\lambda \geq 0}\{-\psi_P(-\lambda) - \psi_Q(\lambda)\} = -\inf_{\lambda \geq 0}\{\psi_Q(\lambda) + \psi_P(-\lambda)\}$$

$$\psi_Q(\lambda) + \psi_P(-\lambda) = \log(\mathbb{E}e^{\lambda W_i}) + \log(\mathbb{E}e^{-\lambda Z_i}) = \log(\mathbb{E}[e^{\lambda(W_i - Z_i)}])$$

$$\mathbb{E}[e^{\lambda W_i}]\mathbb{E}[e^{-\lambda Z_i}] = \left(\sum_x e^{\lambda W_i} q(x)\right)\left(\sum_x e^{-\lambda Z_i} p(x)\right) \geq \left(\sum_x \sqrt{p(x)q(x)}\, e^{\lambda(W_i - Z_i)}\right)^2$$

$$= \left(\sum_x \sqrt{p(x)q(x)}\, \exp\left(\lambda\left[\cancel{\log\left(\frac{p(x)}{q(x)}\right) + \log\left(\frac{p(x)}{q(x)}\right)}\right]\right)\right)^2 = \left(\sum_x \sqrt{p(x)q(x)}\right)^2 = B(P,Q)^2$$

$$\left(\sum e^{\lambda W_i} q(x)\right)\left(\sum e^{-\lambda Z_i} p(x)\right) = \left(\sum_x \left(\frac{p(x)}{q(x)}\right)^{\lambda} q(x)\right)\left(\sum_x \left(\frac{q(x)}{p(x)}\right)^{\lambda} p(x)\right) = \Big(\overbrace{\sum_x p(x)^{\lambda} q(x)^{1-\lambda}}^{b_1}\Big)\Big(\overbrace{\sum_x q(x)^{\lambda} p(x)^{1-\lambda}}^{b_2}\Big)$$

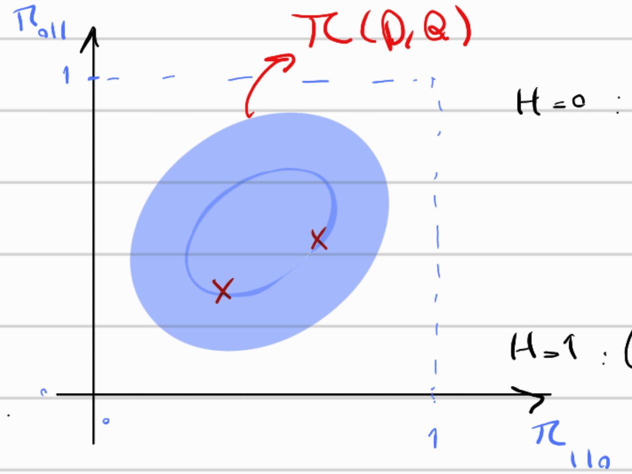
@ $\lambda = \frac{1}{2}$: $b_1 = b_2 = B(P,Q)$ $\Rightarrow$ equality at $\lambda = \frac{1}{2}$

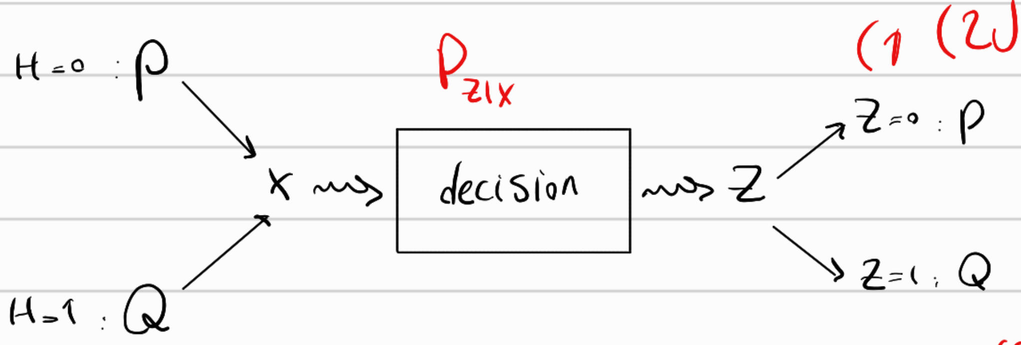
$$\psi_Q(\lambda) + \psi_P(-\lambda) = \log\left(\mathbb{E}(e^{\lambda W_i})\mathbb{E}(e^{-\lambda Z_i})\right) \geq \log(B(P,Q)^2) = -\alpha$$

$\Rightarrow$ equality is inf $\Rightarrow$ inf @ $\lambda = \frac{1}{2}$ $\Rightarrow$ $F_0 = -\inf_{\lambda \geq 0}(\cdots) = \alpha$ $\Rightarrow$ $\boxed{F(0) = \alpha}$

$$F(t) = \sup_{\lambda \geq 0}\{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)\} \geq \{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)\}\Big|_{\lambda = \frac{1}{2}} = \frac{t}{2} + \alpha = \frac{t}{2} + F(0)$$ (3

$$\Rightarrow \boxed{F(t) \geq \frac{t}{2} + F(0)} \xrightarrow{\text{putting it together}} \boxed{\mathbb{P}\left[\sum W_i - Z_i \geq nt\right] \leq \exp(-nF(t)) \leq \exp\left(-n\left(\alpha + \frac{t}{2}\right)\right)}$$





(1 (2الف

(a

$\pi_{1|0} = P(Z=1)$, $\pi_{0|1} = Q(Z=0)$

assume for some test: $P^{I}_{Z|X}$, we have $(\pi^{I}_{1|0}, \pi^{I}_{0|1})$ and for $P^{II}_{Z|X}$: $(\pi^{II}_{1|0}, \pi^{II}_{0|1})$

$$\pi_{1|0} = P(Z=1|H=0) = \sum_x P(x) P_{Z|X}(1|x) \quad , \quad \pi_{0|1} = \sum_x Q(x) P_{Z|X}(0|x) = P(Z=0|H=1)$$

now, let $t \sim Ber(\alpha)$: $P'_{tZ|X} = Ber(\alpha) \times P_{Z|X} \Rightarrow P'_{Z|X} = \alpha P^{I}_{Z|X} + (1-\alpha) P^{II}_{Z|X}$

since equations are linear: $\pi'_{1|0} = \alpha \pi^{I}_{1|0} + (1-\alpha)\pi^{II}_{1|0}$, $\pi'_{0|1} = \alpha \pi^{I}_{0|1} + (1-\alpha)\pi^{II}_{0|1}$

$\Rightarrow$ the convex Hull is achievable $\Rightarrow$ $\boxed{\pi(P,Q) \text{ is convex}}$

(b

Neyman pearson:

$$\text{ideal test: } \left\{ \begin{array}{l} \log\left(\frac{P^{\otimes n}(x)}{Q^{\otimes n}(x)}\right) \geq \tau \to H_0 \\ \log\left(\frac{P^{\otimes n}(x)}{Q^{\otimes n}(x)}\right) < \tau \to H_1 \end{array} \right\} \Rightarrow \begin{array}{l} Q(Z=0) = Q\left(\log\left(\frac{P^{\otimes n}(x)}{Q^{\otimes n}(x)}\right) \geq \tau\right) = \pi_{0|1} \\ P(Z=1) = P\left(\log\left(\frac{P^{\otimes n}(x)}{Q^{\otimes n}(x)}\right) < \tau\right) = \pi_{1|0} \end{array}$$

$$\log\left(\frac{P^{\otimes n}(x)}{Q^{\otimes n}(x)}\right) = \sum_{i=1}^{n} \log\left(\frac{P(x_i)}{Q(x_i)}\right) , \quad \tau = nt \Rightarrow Q(Z=0) = Q\left(\sum_{i=1}^{n} \log\left(\frac{P(x_i)}{Q(x_i)}\right) \geq nt\right)$$

chernoff, similiar to problem 1

$$= P\left[\sum_{i=1}^{n} \log\left(\frac{P(x_i)}{Q(x_i)}\right) \geq nt \,\middle|\, X_i \sim Q\right] \leq \exp\Big\{-n\Big(\lambda t - \underbrace{\log \mathbb{E}_{X_i\sim Q}\Big[\exp\Big(\lambda \log\Big(\frac{P(x)}{Q(x)}\Big)\Big)\Big]}_{\psi_Q(\lambda)}\Big)\Big\}$$

$$= \exp\{-n(\lambda t - \psi_Q(\lambda))\} \quad \forall \lambda \geq 0$$

$$P(Z=1) = P\left[\sum \log\left(\frac{Q(x_i)}{P(x_i)}\right) \geq -nt \,\middle|\, X_i \sim P\right] \leq \exp\Big(n\Big(\lambda t + \underbrace{\log \mathbb{E}_{X_i\sim P} \exp\Big(\lambda \log\Big(\frac{Q(x)}{P(x)}\Big)\Big)}_{\psi_P(\lambda)}\Big)\Big)$$

$$= \exp(-n[-\lambda t - \psi_P(\lambda)]) \quad \forall \lambda \geq 0$$

$$\Rightarrow Q(Z=0) = \pi_{0|1} \leq \inf_{\lambda} \exp(-n(\lambda t - \psi_Q(\lambda))) = \exp\Big(-n \underbrace{\sup_{\lambda \geq 0}\{\lambda t - \psi_Q(\lambda)\}}_{\psi_Q^*(t)}\Big)$$

$\Rightarrow$ $\boxed{\pi_{0|1} \leq \exp(-n\psi_Q^*(t))}$, same process for $P(Z=1)$

$\Rightarrow$ similarly : $\boxed{P(Z=1) \leq \exp(-n \psi_P^*(t))}$ (our $\sup_{\lambda \geq 0}$ is not the same, argued later)

conditions? the sup should converge

$$\psi_P(\lambda) = \log E_{x\sim p} \exp(\lambda \log(\tfrac{q(x)}{p(x)})) \geq E_{x\sim p} \log \exp(\lambda \log(\tfrac{q(x)}{p(x)})) = \lambda \mathbb{E}_{x\sim p} \log(\tfrac{q(x)}{p(x)})$$

$$\psi_P(\lambda) \geq -\lambda D_{KL}(P\|Q) \overset{\text{similiar}}{\rightsquigarrow} \psi_Q(\lambda) \geq -\lambda D_{KL}(Q\|P)$$

$$\Rightarrow -\lambda t - \psi_P(\lambda) \leq -\lambda(t - D_{KL}(P\|Q)) \Rightarrow \psi_P^*(t) \leq \sup_{\lambda \geq 0}\{-\lambda(t - D_{KL}(P\|Q))\} = \begin{cases} 0 & t \geq D_{KL}(P\|Q) \\ \infty & \text{o.w.} \end{cases}$$

$$, \lambda t - \psi_Q(\lambda) \leq \lambda(t + D_{KL}(Q\|P)) \Rightarrow \psi_Q^*(t) \leq \sup_{\lambda \geq 0}\{\lambda(t + D_{KL}(Q\|P))\} = \begin{cases} 0 & t \leq -D_{KL}(Q\|P) \\ \infty & \text{o.w.} \end{cases}$$

$\boxed{\text{if } -D_{KL}(Q\|P) \leq t \leq D_{KL}(P\|Q) \Rightarrow \sup_{\lambda \geq 0} \text{ can become } \sup_{\lambda \in \mathbb{R}}}$

now, after replacing with $\lambda \in \mathbb{R}$, if t is outside this region, ψ_P^*, ψ_Q^* diverge.

using lose 2: $\boxed{\pi_{1|0}^{(n)} \leq 2^{-n\psi_P^*(t)}, \ \pi_{0|1}^{(n)} \leq 2^{-n\psi_Q^*(t)}, \text{ if } -D_{KL}(Q\|P) \leq t \leq D_{KL}(P\|Q)}$

(3

$$\forall \lambda : \psi_Q(\lambda) = \log E_{x\sim Q}[e^{\lambda \log(\frac{p(x)}{q(x)})}] = \log \mathbb{E}_{x\sim Q}[(\tfrac{p(x)}{q(x)})^\lambda] = \log \sum_{x \in \mathcal{X}} p(x)^\lambda q(x)^{1-\lambda}$$

$$\psi_P(1-\lambda) = \log \mathbb{E}_{x\sim p}[e^{(1-\lambda)\log(\frac{q(x)}{p(x)})}] = \log E_{x\sim p}[(\tfrac{q(x)}{p(x)})^{1-\lambda}] = \log \sum_{x\in\mathcal{X}} p(x)^\lambda q(x)^{1-\lambda}$$

$$\Rightarrow \psi_Q(\lambda) = \psi_P(1-\lambda)$$

$$\psi_Q^*(t) = \sup_{\lambda \in \mathbb{R}}\{\lambda t - \psi_Q(\lambda)\} = \sup_{\lambda \in \mathbb{R}}\{\lambda t - \psi_P(1-\lambda)\} = \sup_{\gamma \in \mathbb{R}}\{(1-\gamma)t - \psi_P(\gamma)\}$$

$$= t + \sup_{\gamma \in \mathbb{R}}\{-\gamma t - \psi_P(\gamma)\} = t + \psi_P^*(t) = \psi_Q^*(t) = E_1(t), \ E_0(t) = \psi_P^*(t)$$

$$\Rightarrow \text{if } -D_{KL}(Q\|P) \leq t \leq D_{KL}(P\|Q) \Rightarrow \pi_{1|0}^{(n)} \leq e^{-nE_0} \leq e^{-n\psi_P^*(t)} \Rightarrow \psi_P^*(t) < E_0$$

$$, \pi_{0|1}^{(n)} \leq e^{-nE_1} \leq e^{-n\psi_Q^*(t)} \Rightarrow \psi_Q^*(t) = t + \psi_P^*(t) \leq E_1$$

the boundaries are $E_0, E_1 \Rightarrow \leq E_0, E_1$ is achievable (ψ^*)

$\Rightarrow \boxed{\text{if } -D_{KL}(Q\|P) \leq t \leq D_{KL}(P\|Q) : E_0(t) = \psi_P^*(t), E_1(t) = \psi_P^*(t) + t \text{ : achievable}}$

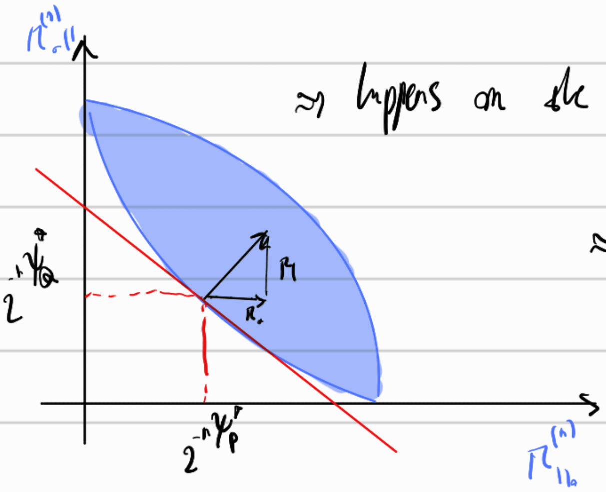
by Sanov's theorem, as shown in class we know that (4

for $n \to \infty$, the chernoff bound is tight, thus giving us the optimum possible exponent (up to log factors) $\rightsquigarrow$ for $n \to \infty$, the bound is tight,

for $n < \infty$: $E_0(t) \leq E_0$, $E_1(t) \leq E_1$ $\rightsquigarrow$ $\boxed{\text{boundary}: E_0(t) = \psi_P^*(t),\ E_1(t) = \psi_P^*(t) + t}$

(5

$$\min_{P(Z|X^n)} \{\pi_0 \pi_{1|0} + \pi_1 \pi_{0|1}\}$$



$\rightsquigarrow$ happens on the edge case $\rightsquigarrow$ $\pi_{1|0} = 2^{-nE_0}$, $\pi_{0|1} = 2^{-nE_1}$

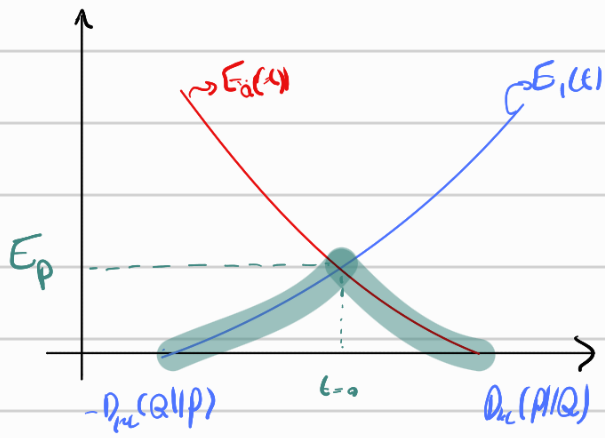
$$\rightsquigarrow P_E = \pi_0 2^{-nE_0} + \pi_1 2^{-nE_1} \leq 2^{-nE_P}$$

at $n \to \infty$: the LHS exponent will be the smaller one

$\Rightarrow \min\{E_0, E_1\} \leq E_P \longrightarrow$ equality occurs on border $\rightsquigarrow$ maximum exponent

$\longrightarrow P_E \leq 2^{-nE_P}$, $E_P = \max_{(E_0,E_1) \in \mathcal{E}} \min\{E_0, E_1\}$, edge is $E_0(t), E_1(t)$

$$\rightarrow \boxed{E_P = \max_{(E_0,E_1) \in \mathcal{E}} \min\{E_0, E_1\} = \max_{t \in \mathcal{T}} \min\{E_0(t), E_1(t)\},\ \mathcal{T} = [-D_{KL}(Q\|P), D_{KL}(P\|Q)]}$$



$\rightsquigarrow$ maxmin occurs when $E_0(t) = E_1(t)$

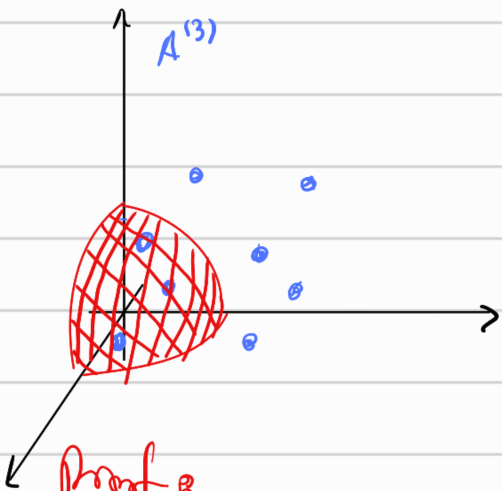
$$\Rightarrow \psi_P^*(t) + t = \psi_P^*(t) \Rightarrow \boxed{t = 0}$$

$$\Rightarrow \boxed{E_P = \psi_P^*(0)}$$

$$\psi_P^*(0) = \sup_{\lambda \in \mathbb{R}} \{-\psi_P(\lambda)\} = -\inf_{\lambda \in \mathbb{R}} \{\psi_P(\lambda)\}$$ (later in Q4)

سوال (3)

$B^{(n)} = \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$, $A^{(n)} = \{x \in \mathbb{R}^n : \forall i \in [n]\ x_i \in \{1,2\}\}$



$$f(r_0, n) = \sum_{a \in A^{(n)}} \mathbb{1}\{a \notin B^{(n)}(r_0\sqrt{n})\}$$

theorem:

$\exists r^* \text{ st. } \quad e[f(r_0,n)] = 1 \quad \forall r_0 < r^*$

$e[f(r_0,n)] < 1 \quad \forall r_0 > r^*$

$e[h(n)] = \lim_{n\to\infty} \frac{\log(h(n))}{n}$

Proof B

consider the probability, uniformly choosing $a \sim \text{Unif}\{A^{(n)}\}$

$$\mathbb{P}\{a \notin B^{(n)}(r_0\sqrt{n})\} = \frac{N_{in}}{N_{tot}} = \frac{f(r_0,n)}{|A^{(n)}|} , \quad |A^{(n)}| = 2^n$$

$$\rightarrow f(r_0,n) = 2^n \mathbb{P}\{a \notin B^{(n)}(r_0\sqrt{n})\} \approx a 2^{nE} \rightsquigarrow E \leq 1 \text{ obviously}$$

$$\mathbb{P}\{a \notin B^{(n)}(r_0\sqrt{n})\} = \mathbb{P}\{|x| \geq r_0\sqrt{n} \mid x \sim \text{Unif}\{A^{(n)}\}\} = \mathbb{P}\{\|x\|^2 \geq n r_0^2 \mid x \sim \text{Unif}\{A^{(n)}\}\}$$

$$= \mathbb{P}\{\sum x_i^2 \geq n r_0^2 \mid x_i \sim \text{Unif}(\{1,2\})\} = \mathbb{P}\{\sum w_i \geq n r_0^2 \mid w_i \sim \text{Unif}(\{1,4\})\}$$

$$= \mathbb{P}\{\tfrac{1}{n}\sum w_i \geq r_0^2 \mid w_i \sim \text{Unif}(\{1,4\})\} ; \quad \bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} w_i$$

$\mathbb{E}\bar{W}_n = \mathbb{E}W_i = \frac{1}{2}\times 1 + \frac{1}{2}\times 4 = 2.5$; by Law of Large numbers: $\forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}[\bar{W}_n > \mathbb{E}\bar{W}_n - \epsilon] = 1$

$\Rightarrow$ if $r_0^2 < \mathbb{E}\bar{W}_n = 2.5 \Rightarrow \lim_{n\to\infty} \mathbb{P}\{\bar{W}_n > r_0^2\} = 1 \rightarrow f(r_0,n) \to 2^n \Rightarrow \boxed{E = 1 : r_0^2 < 2.5}$

o.w. $\rightsquigarrow \mathbb{P}\{\sum w_i^2 \geq n r_0^2\} \leq 2^{-\lambda n r_0^2}\, \mathbb{E}[2^{\lambda \sum w_i^2}] = \exp_2\{-n(\lambda r_0^2 - \log \mathbb{E}\, 2^{\lambda w_i})\}$; $\mathbb{E}2^{\lambda w_i} = \frac{2^\lambda + 2^{4\lambda}}{2}$

$\rightarrow f(r_0,n) \leq \exp_2\{-n(\lambda r_0^2 - \log(2^\lambda + 2^{4\lambda}))\} \rightsquigarrow E \leq \lambda r_0^2 - \log(2^\lambda + 2^{4\lambda})\ \forall \lambda$

$\rightarrow E \leq \inf_{\lambda \geq 0}\{\log(2^\lambda + 2^{4\lambda}) - \lambda r_0^2\} \rightsquigarrow \frac{d}{d\lambda} : -r_0^2 + 1 + \frac{3 \cdot 2^{3\lambda}}{1 + 2^{3\lambda}}$: monatonic $\rightsquigarrow$ if $\frac{d}{d\lambda}\Big|_0 < 0 \Rightarrow \inf_{\lambda \geq 0} < f(0) = 1$

$\frac{d}{d\lambda}\Big|_0 = \frac{5}{2} - r_0^2 \Rightarrow \boxed{\text{if } r_0^2 > 2.5 \Rightarrow E < 1} \rightarrow \boxed{r^* = \sqrt{2.5}}$

سوال ۴)

$X_1, X_2, \dots, X_n \sim Q$, $\pi_1 : Q = P_1$, $\pi_2 : Q = P_2$: $A^{(n)} \subseteq \mathcal{X}^n$: H_1

$\alpha_n = P_1^{(n)}(\mathcal{X}^n \setminus A^{(n)})$, $\beta_n = P_2^{(n)}(A^{(n)})$, $P_E^{(n)} = \pi_1 \alpha_n + \pi_2 \beta_n$, $D^* = \lim_{n\to\infty} \frac{-1}{n} \log \min_{A^{(n)} \in \mathcal{X}^n} \{P_E^{(n)}\}$

1)

$$P_E^{(n)} = \pi_1 \alpha_n + \pi_2 \beta_n = \pi_1 P_1^{(n)}(\mathcal{X}^n \setminus A^{(n)}) + \pi_2 P_2^{(n)}(A^{(n)}) = \pi_1 + \pi_2 P_2^{(n)}(A^{(n)}) - \pi_1 P_1^{(n)}(A^{(n)})$$

$$\Rightarrow \min_{A^{(n)} \in \mathcal{X}^n} P_E^{(n)} = \min_{A^{(n)} \in \mathcal{X}^n} \{\pi_1 + \pi_2 P_2^{(n)}(A^{(n)}) - \pi_1 P_1^{(n)}(A^{(n)})\} = \pi_1 + \min_{A^{(n)} \in \mathcal{X}^n} \{\pi_2 P_2^{(n)}(A^{(n)}) - \pi_1 P_1^{(n)}(A^{(n)})\}$$

$$= \pi_1 + \min_{A^{(n)} \in \mathcal{X}^n} \Big\{ \sum_{x^n \in \mathcal{X}^n} (\pi_2 P_2^{(n)}(x^n) - \pi_1 P_1^{(n)}(x^n)) \mathbb{1}(x^n \in A^{(n)}) \Big\} \xrightarrow{\min} A^{*(n)} = \{x^n \mid \pi_1 P_1^{(n)}(x^n) \geq \pi_2 P_2^{(n)}(x^n)\}$$

$$\Rightarrow \min_{A^{(n)} \in \mathcal{X}^n} P_E^{(n)} = \pi_1 - \sum_{x^n \in \mathcal{X}^n} [\pi_1 P_1^{(n)}(x^n) - \pi_2 P_2^{(n)}(x^n)]_+ = \pi_1 - \pi_1 P_1^{(n)}\Big[\frac{P_1^{(n)}(x^n)}{P_2^{(n)}(x^n)} > \frac{\pi_2}{\pi_1}\Big] + \pi_2 P_2^{(n)}\Big[\frac{P_1^{(n)}(x^n)}{P_2^{(n)}(x^n)} > \frac{\pi_2}{\pi_1}\Big]$$

$$\Rightarrow \min_{A^{(n)} \in \mathcal{X}^n} P_E^{(n)} = \pi_1 P_1^{(n)}\Big[\frac{P_1^{(n)}(x^n)}{P_2^{(n)}(x^n)} < \frac{\pi_2}{\pi_1}\Big] + \pi_2 P_2^{(n)}\Big[\frac{P_1^{(n)}(x^n)}{P_2^{(n)}(x^n)} > \frac{\pi_2}{\pi_1}\Big] \rightsquigarrow \text{same as } 2$$

from 2: $E_P = \psi_{P_1}^*(0) = \psi_{P_2}^*(0) = \sup_{\lambda \in \mathbb{R}} \{-\psi_{P_2}(\lambda)\} = -\inf_{\lambda \in \mathbb{R}} \{\psi_{P_2}(\lambda)\} = -\psi_{P_2}(\lambda^*)$

$$\psi_{P_2}(\lambda) = \log E_{x\sim P_2}\Big[\exp\Big(\lambda \log\Big(\frac{P_1(x)}{P_2(x)}\Big)\Big)\Big] \rightsquigarrow \inf_{\lambda \in \mathbb{R}} E_{x\sim P_2}\Big[\exp\Big(\lambda \log\Big(\frac{P_1(x)}{P_2(x)}\Big)\Big)\Big]$$

$$\Rightarrow \frac{d}{d\lambda} = 0 \Rightarrow E_{x\sim P_2}\Big[\log\Big(\frac{P_1(x)}{P_2(x)}\Big)\exp\Big(\lambda^* \log\Big(\frac{P_1(x)}{P_2(x)}\Big)\Big)\Big] = 0 \Rightarrow E_{x\sim P_2}\Big[\Big(\frac{P_1(x)}{P_2(x)}\Big)^{\lambda^*}\log\Big(\frac{P_1(x)}{P_2(x)}\Big)\Big]$$

$$\Rightarrow \sum_{x\in\mathcal{X}} P_1^{\lambda^*}(x) P_2^{1-\lambda^*}(x) \log\Big(\frac{P_1(x)}{P_2(x)}\Big) = 0 \rightsquigarrow \frac{1}{\sum_{x\in\mathcal{X}} P_1^{\lambda^*}(x) P_2^{1-\lambda^*}(x)} \sum_{x\in\mathcal{X}} \underbrace{P_1^{\lambda^*}(x) P_2^{1-\lambda^*}(x)}_{\to P_\lambda} \log\Big(\frac{P_1(x)}{P_2(x)}\Big) = 0$$

$\Rightarrow$ define : $P_\lambda(x) = \dfrac{P_1^\lambda(x) P_2^{1-\lambda}(x)}{\sum_{y\in\mathcal{X}} P_1^\lambda(y) P_2^{1-\lambda}(y)}$, $f(\lambda) = E_{P_\lambda}\Big[\log\Big(\frac{P_1(x)}{P_2(x)}\Big)\Big]$, we need : $f(\lambda^*) = 0$

obviously : $f(\lambda)$ is continuous ; $f(0) = E_{P_2}\Big[\log\Big(\frac{P_2(x)}{P_1(x)}\Big)\Big] = D_{KL}(P_2 \| P_1) \geq 0$

$f(1) = E_{P_1}\Big[\log\Big(\frac{P_2(x)}{P_1(x)}\Big)\Big] = -D_{KL}(P_1 \| P_2) \leq 0 \Rightarrow f(\lambda) : \text{cont}, f(1) \leq 0, f(0) \geq 0 \Rightarrow \boxed{\exists \lambda^* \in [0,1] : f(\lambda^*) = 0}$

$$\Rightarrow D^* = -\psi_{P_1}^*(\lambda^*) = -\psi_{P_2}^*(\lambda^*) \quad , \exists \lambda^* \in [0,1], @ \lambda^* : f(\lambda^*) = 0$$

now, we find its relation with D_{KL}.

$$D^* = -\psi^*_{P_2}(\lambda^*) = -\log \mathbb{E}_{\lambda\sim P_2}\left[\exp\left(\lambda^*\log\left(\frac{P_1(x)}{P_2(x)}\right)\right)\right] = -\log \mathbb{E}_{x\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda^*}\right]$$

$$= -\log\left(\sum_{\lambda\in\mathcal{X}} P_1^{\lambda^*}(x)P_2^{1-\lambda^*}(x)\right) = \lambda^* \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\frac{P_1(x)}{P_2(x)}\right)\right] - \log\left(\sum_{y\in\mathcal{X}} P_1^{\lambda^*}(y)P_2^{1-\lambda^*}(y)\right)$$

$$= \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\frac{P_1^{\lambda^*}(x)}{P_2^{\lambda^*}(x)}\right) - \log\left[\sum_{y\in\mathcal{X}} P_1^{\lambda^*}(x)P_2^{1-\lambda^*}(x)\right]\right] = \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\frac{P_1^{\lambda^*}(x)P_2^{-\lambda^*}(x)}{\sum_{y\in\mathcal{X}}P_1^{\lambda^*}(y)P_2^{1-\lambda^*}(x)}\right)\right]$$

$$= \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\frac{P_{\lambda^*}(x)}{P_2(x)}\right)\right] = D_{KL}(P_{\lambda^*}\|P_2)$$

$$D_{KL}(P_{\lambda^*}\|P_1) = \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\frac{P_{\lambda^*}(x)}{P_1(x)}\right)\right] = \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\frac{P_{\lambda^*}(x)}{P_2(x)}\right) - \cancelto{0}{\log\left(\frac{P_1(x)}{P_2(x)}\right)}\right] = D_{KL}(P_{\lambda^*}\|P_2)$$

$$\Rightarrow \boxed{D^* = D_{KL}(P_{\lambda^*}\|P_1) = D_{KL}(P_{\lambda^*}\|P_2)}$$

(b)

$$D_{KL}(P_{\lambda^*}\|P_2) = \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\frac{P_{\lambda^*}(x)}{P_2(x)}\right)\right] = \mathbb{E}_{P_{\lambda^*}}\left[\log\left(\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\frac{1}{\sum_{y\in\mathcal{X}}P_1^{\lambda}(y)P_2^{1-\lambda}(y)}\right)\right]$$

$$= \mathbb{E}_{P_{\lambda^*}}\left[\lambda\log\left(\frac{P_1(x)}{P_2(x)}\right) - \log\Bigg(\underbrace{\sum_{y\in\mathcal{X}}P_2(y)\left(\frac{P_1(y)}{P_2(y)}\right)^{\lambda}}_{\mathbb{E}_{P_2}\left[\left(\frac{P_1(y)}{P_2(y)}\right)^{\lambda}\right]}\Bigg)\right] = \lambda\,\mathbb{E}_{P_{\lambda^*}}\left[\cancelto{0}{\log\left(\frac{P_2(x)}{P_1(x)}\right)}\right] - \log \mathbb{E}_{P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda^*}\right]$$

$$\Rightarrow D^* = -\log\left(\mathbb{E}_{x\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda^*}\right]\right), \quad \text{let: } C(\lambda) = -\log \mathbb{E}_{x\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\right]$$

$$\Rightarrow C(\lambda) = -\log \mathbb{E}_{x\sim P_2}\left[\exp\left(\lambda\log\left(\frac{P_1(x)}{P_2(x)}\right)\right)\right] : \frac{dC(\lambda)}{d\lambda} = -\frac{\mathbb{E}_{x\sim P_2}\left[\log\left(\frac{P_1(x)}{P_2(x)}\right)\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\right]}{\mathbb{E}_{x\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\right]}$$

$$\Rightarrow \frac{dC(\lambda)}{d\lambda} = -\frac{\sum \log\left(\frac{P_1(x)}{P_2(x)}\right)P_1^{\lambda}(x)P_2^{1-\lambda}(x)}{\sum_{x\in\mathcal{X}}P_1^{\lambda}(x)P_2^{1-\lambda}(x)} = -\sum P_{\lambda}(x)\log\left(\frac{P_1(x)}{P_2(x)}\right) = f(\lambda) \quad \Rightarrow \frac{dC}{d\lambda}\bigg|_{\lambda^*} = 0$$

$\Rightarrow$ optimal at λ^* $\Rightarrow$ $\boxed{D^* = -\log\left(\mathbb{E}_{x\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda^*}\right]\right) = -\min_{0\leq\lambda\leq 1}\log\left(\mathbb{E}_{x\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\right]\right)}$

$f(0)\geq 0,\ f(1)\leq 0 \Rightarrow C(\lambda)$:

$\Rightarrow$ $\boxed{\lambda^* \text{ is the maximum.}}$

$$\Rightarrow D^* = \max_{0\leq\lambda\leq 1}\left\{-\log\left(\mathbb{E}_{x\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\right]\right)\right\}$$

$X \sim \text{Binomial}(n,p)$ (سوال ۲

(1

let: $X_i \sim \text{Bern}(p) \Rightarrow X = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n,p)$

$$\mathbb{P}\left\{\begin{matrix} X \geq k \\ X \leq k \end{matrix}\right\} = \mathbb{P}\left\{\begin{matrix} \sum_{i=1}^{n} X_i \geq k \\ \sum_{i=1}^{n} X_i \leq k \end{matrix}\right\} \underset{\lambda \leq 0}{\overset{\lambda \geq 0}{=}} \mathbb{P}\{e^{\lambda \sum X_i} \geq e^{\lambda k}\} \leq e^{-\lambda k} E[e^{\lambda \sum X_i}] = e^{-\lambda k} E[e^{\lambda X_i}]^n$$

$$= \exp\left(-\lambda k + n \log E[e^{\lambda X_i}]\right) = \exp\left(-n\left(\lambda \tfrac{k}{n} - \log E[e^{\lambda X_i}]\right)\right) \; ; \; E[e^{\lambda X_i}] = pe^{\lambda} + (1-p)$$

$$\Rightarrow \mathbb{P}\{X \geq k\} \leq \exp\left(-n\left[\lambda \tfrac{k}{n} - \log(1 + p(e^{\lambda} - 1))\right]\right) \quad \lambda \geq 0, \; \mathbb{P}\{X \leq k\} \text{ - same but } \lambda \leq 0$$

$$\Rightarrow \mathbb{P}\{X \geq k\} \leq \inf_{\lambda \geq 0}\left\{\exp\left(-n\left[\lambda \tfrac{k}{n} - \log(1 + p(e^{\lambda}-1))\right]\right)\right\} = \exp\left(-n \sup_{\lambda \geq 0}\left\{\lambda \tfrac{k}{n} - \log(1 + p(e^{\lambda}-1))\right\}\right)$$

$$\frac{d}{dx}\left(\lambda \tfrac{k}{n} - \log(1 - p + pe^{\lambda})\right) = \frac{k}{n} - \frac{pe^{\lambda}}{1 - p + pe^{\lambda}} = 0 \Rightarrow \frac{n}{k} = \frac{1-p}{p} e^{-\lambda} + 1 \Rightarrow \frac{n}{k} - 1 = \left(\frac{1}{p} - 1\right) e^{-\lambda}$$

$$\Rightarrow \lambda^* = \ln\left(\frac{\frac{k}{n}(1-p)}{p(1-\frac{k}{n})}\right) = \ln\left(\frac{1-p}{1-\frac{k}{n}}\right) - \ln\left(\frac{p}{\frac{k}{n}}\right), \quad \log(1 - p + pe^{\lambda}) = \log\left(\frac{pe^{\lambda}}{\frac{k}{n}}\right) = \lambda^* + \log\left(\frac{p}{\frac{k}{n}}\right)$$

$\lambda \geq 0 \Rightarrow \lambda^{opt} = \lambda^* \wedge 0 \Rightarrow$ the 0 bound gives $\mathbb{P}\{\dots\} \leq 1$ which is obv.

$$\sup\{\dots\} = \frac{k}{n} \log\left(\frac{\frac{k}{n}}{p}\right) + \left(1 - \frac{k}{n}\right) \log\left(\frac{1 - \frac{k}{n}}{1-p}\right) = D_{KL}\left(\text{Bern}\left(\tfrac{k}{n}\right) \| \text{Bern}(p)\right) = d_{KL}\left(\tfrac{k}{n} \| p\right)$$

$\frac{k}{n} > p: \lambda^* > 0, \quad \frac{k}{n} < p: \lambda^* < 0 \Rightarrow$ bounds for $\geq$ or $\leq$

$$\boxed{\mathbb{P}\{X \geq k\} \leq \exp\{-n\lambda^*\} \;\forall k > np \;;\; \mathbb{P}\{X \leq k\} \leq \exp\{-n\lambda^*\} \;\forall k < np \;;\; \lambda^* = d_{KL}\left(\tfrac{k}{n} \| p\right)}$$

(2

let: $k = uE[X] = unp \Rightarrow k < np \equiv u < 1, \; \frac{k}{n} = up \Rightarrow$ substitute in part 1.

$$\to \mathbb{P}\{X \geq uE[X]\} \leq \exp\{-n d_{KL}(up\|p)\} \;\forall u > 1, \; \mathbb{P}\{X \leq uE[X]\} \leq \exp\{-n d_{KL}(up\|p)\} \; 0 < u < 1$$

now: $$d_{KL}(up\|p) = up\ln(u) + \overset{x}{(1-up)} \ln\left(\frac{\overset{x}{1-up}}{\underset{y}{1-p}}\right) \geq up\ln(u) + [(1-up) - (1-p)] = up\ln(u) + p(1-u)$$

$$\Rightarrow d_{KL}(up\|p) \geq p[u\ln(u) - (u-1)] = pf(u), \quad np = E[X]$$

$$\Rightarrow n d_{KL}(up\|p) \geq np f(u) = E[X] f(u) \Rightarrow \exp\{-n d_{KL}(up\|p)\} \leq \exp\{-E[X] f(u)\}$$

$\Rightarrow$ putting it together $\Longrightarrow \begin{cases} \boxed{\mathbb{P}\{X \geq u E[X]\} \leq \exp\{-E[X] f(u)\} \ \forall u > 1} \\ \boxed{\mathbb{P}\{X \leq u E[X]\} \leq \exp\{-E[X] f(u)\} \ \forall\, 0 \leq u \leq 1} \end{cases}$

$$f(u) = u\ln(u) - (u-1) = u\int_1^u \frac{dx}{x} - \int_1^u dx = \int_1^u \frac{u-x}{x}\, dx \geq \int_1^u \frac{u-x}{u}\, dx = \frac{1}{u}\cdot\frac{(u-1)^2}{2} \Rightarrow \boxed{f(u) \leq \frac{(u-1)^2}{2u}} \quad (3$$

$$\mathbb{P}\{X \geq u E[X]\} \leq \exp\{-E[X] f(u)\} \leq \exp\left\{-E[X]\frac{(u-1)^2}{2u}\right\} \ : \forall u > 1 \ ; \text{ let: } u_0 = \frac{E[X]+t}{E[X]} \quad t > 0$$

$$\Rightarrow \mathbb{P}\{X \geq t + E[X]\} \leq \exp\left\{-\cancel{E[X]}\cdot\frac{\left(\frac{t}{E[X]}\right)^2}{2\times(E[X]+t)/\cancel{E[X]}}\right\} = \exp\left\{-\frac{t^2}{2(E[X]+t)}\right\} \ ; \ np = E[X]$$

$$\Rightarrow \boxed{\mathbb{P}\{X \geq t + np\} \leq \exp\left\{-\frac{t^2}{2(np+t)}\right\} \ : \forall t > 0} \quad (4$$

assuming : $d_{KL}(\frac{k}{n} \| p) \geq \left(\sqrt{\frac{k}{n}} - \sqrt{p}\right)^2$; $k > np \Rightarrow$ let $\sqrt{k} = \sqrt{np} + t$, $k < np \Rightarrow \sqrt{k} = \sqrt{np} - t$ ($t > 0$)

$$\Rightarrow \mathbb{P}\{X \geq k\} = \mathbb{P}\{X \geq (\sqrt{np}+t)^2\} \leq \exp\{-n d_{KL}(\tfrac{k}{n}\|p)\} \leq \exp\left\{-n\left(\frac{t}{\sqrt{n}}\right)^2\right\}$$

$$, X \geq 0 \Rightarrow \mathbb{P}\{X \geq (\sqrt{np}+t)^2\} = \mathbb{P}\{\sqrt{X} \geq \sqrt{np}+t\} = \mathbb{P}\{\sqrt{X} - \sqrt{np} \geq t\} \leq e^{-t^2}$$

similarly : $\sqrt{k} = \sqrt{np} - t$, $0 < t < \sqrt{np} \Rightarrow \mathbb{P}\{X \leq k\} \leq \exp\{-n d_{KL}(\frac{k}{n}\|p)\}$

$\Rightarrow \mathbb{P}\{\sqrt{X} \leq \sqrt{k}\} = \mathbb{P}\{\sqrt{X} \leq \sqrt{np} - t\} \leq \exp(-t^2)$ $\forall\, 0 < t < \sqrt{np}$, for $t > \sqrt{np}$: the result is obvious $\Rightarrow$ holds for $t > 0$ $\rightarrow \mathbb{P}\{\sqrt{X} - \sqrt{np} \leq -t\} \leq e^{-t^2}$

$$\Rightarrow \text{if } \underline{d_{KL}(q\|p) \geq (\sqrt{q}-\sqrt{p})^2} \Rightarrow \boxed{\begin{cases} \mathbb{P}\{\sqrt{X} - \sqrt{np} \geq t\} \leq e^{-t^2} \\ \mathbb{P}\{\sqrt{X} - \sqrt{np} \leq -t\} \leq e^{-t^2} \end{cases} \ \forall t > 0}$$

$d_{KL}(q\|p) \geq (\sqrt{q}-\sqrt{p})^2$:

$$d_{KL}(q\|p) = q\ln\left(\frac{q}{p}\right) + \underbrace{(1-q)\ln\left(\frac{1-q}{1-p}\right)}_{\geq (1-q)-(1-p)} \geq q\ln\left(\frac{q}{p}\right) + p - q = 2\sqrt{q}\left[\sqrt{q}\,\underbrace{\ln\left(\frac{\sqrt{q}}{\sqrt{p}}\right)}_{\geq \sqrt{q}-\sqrt{p}}\right] + p - q$$

$$\geq 2\sqrt{q}\left[\sqrt{q} - \sqrt{p}\right] + p - q = 2q - 2\sqrt{pq} + p - q = q + p - 2\sqrt{pq} = (\sqrt{q}-\sqrt{p})^2$$

$$\Rightarrow \boxed{d_{KL}(q\|p) \geq (\sqrt{q}-\sqrt{p})^2}$$

سوال 6)

$$V_{\alpha,\beta,r,s,t}(P_x\|Q_x) = \sup_{f:\mathcal{X}\to\mathbb{R}} \Big\{ \mathbb{E}_{P_x}[f(x)] - r\,\mathbb{E}_{Q_x}[f(x)] - s\log\mathbb{E}_{Q_x}[\exp(\alpha f(x))] - t\log\mathbb{E}_{Q_x}[\exp(\beta f(x))] \Big\}$$

$\alpha,\beta,r\in\mathbb{R},\ s,t\in\mathbb{R}^{++}$

1)

$$V_{\alpha,\beta,r,s,t}(P_x\|Q_x) = \sup_{f:X\to\mathbb{R}} \Big\{ \underbrace{\mathbb{E}_{P_x}[f(x)] - r\,\mathbb{E}_{Q_x}[f(x)] - s\log\mathbb{E}_{Q_x}[e^{\alpha f(x)}] - t\log\mathbb{E}_{Q_x}[e^{\beta f(x)}]}_{\mathcal{L}_{\alpha,\beta,r,s,t}(f)} \Big\}$$

move along adding constants to f, assuming f^* is optimal

$$\mathcal{L}_{\alpha,\beta,r,s,t}(f^*+c) = \mathbb{E}_{P_x}[f^*(x)+c] - r\,\mathbb{E}_{Q_x}[f^*(x)+c] - s\log\mathbb{E}_{Q_x}[e^{\alpha f^*(x)}e^{\alpha c}] - t\log\mathbb{E}_{Q_x}[e^{\beta f^*(x)}e^{\beta c}]$$

$$= \mathcal{L}_{\alpha,\beta,r,s,t}(f^*) + c\,(\underbrace{1-r-\alpha s-\beta t}_{\Delta})$$

if $\Delta\neq 0$, we can increase $\mathcal{L}(f)$ by adding constant $\Rightarrow \sup\mathcal{L}(f)\to\infty$

$$\Rightarrow \Delta = 0 \Rightarrow \boxed{1 = r+\alpha s+\beta t}$$

2)

$$V_{\alpha,\beta,r,s,t}(P_x\|Q_x) = \sup_{f:X\to\mathbb{R}} \Big\{ \mathbb{E}_{P_x}[f(x)] - r\,\mathbb{E}_{Q_x}[f(x)] - s\log\mathbb{E}_{Q_x}[e^{\alpha f(x)}] - t\log\mathbb{E}_{Q_x}[e^{\beta f(x)}] \Big\}$$

$$\geq \Big\{ \mathbb{E}_{P_x}[f(x)] - r\,\mathbb{E}_{Q_x}[f(x)] - s\log\mathbb{E}_{Q_x}[e^{\alpha f(x)}] - t\log\mathbb{E}_{Q_x}[e^{\beta f(x)}] \Big\}\Big|_{f(x)=c}$$

$$= c - rc - s\log(e^{\alpha c}) - t\log(e^{\beta c}) = c(1-r-\alpha s-\beta t) = 0 \Rightarrow \boxed{V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\geq 0}$$

3)

$$V_{\alpha,\beta,r,s,t}(P_x\|P_x) = \sup_{f:X\to\mathbb{R}} \Big\{ \mathbb{E}_{P_x}[f(x)] - r\,\mathbb{E}_{P_x}[f(x)] - s\log\mathbb{E}_{P_x}[e^{\alpha f(x)}] - t\log\mathbb{E}_{P_x}[e^{\beta f(x)}] \Big\}$$

$$\leq \sup_{f:X\to\mathbb{R}} \Big\{ \mathbb{E}_{P_x}[f(x)] - r\,\mathbb{E}[f(x)] - s\log(\underbrace{e^{\mathbb{E}_{P_x}[\alpha f(x)]}}_{\alpha\mathbb{E}[f(x)]}) - t\log(\underbrace{e^{\mathbb{E}_{P_x}[\beta f(x)]}}_{\beta\mathbb{E}[f(x)]}) \Big\}$$

$$= \sup_{f:X\to\mathbb{R}} \Big\{ (\underbrace{1-r-\alpha s-\beta t}_{0})\,\mathbb{E}_{P_x}[f(x)] \Big\} = 0 \Rightarrow 0\leq V_{\alpha,\beta,r,s,t}(P_x\|P_x)\leq 0 \Rightarrow \boxed{V_{\alpha,\beta,r,s,t}(P_x\|P_x)=0}$$

is this both ways? if $\exists Q_x : V_{\alpha,\beta,r,s,t}(P_x\|Q_x)=0$, it means Yensen has resulted in equality $\Rightarrow \mathbb{E}[f(x)] = f(\mathbb{E}X)$, on convex $f \Rightarrow X$ is constant $\Rightarrow f(X)$ is constant $\Rightarrow f$ is constant $\Rightarrow f(x)=c$, alternatively calculus of variation: $\partial_f \mathcal{L}_{\alpha,\beta,r,s,t}(f) = 0 \,\big|_{f=c}$ from part 2 same result.

$\Rightarrow$ variation: (also note that $\mathcal{L}_{\alpha,\beta,r,s,t}$ is strictly concave $\Rightarrow$ only $f=c$)

$$\mathcal{L}_{\alpha,\beta,r,s,t}(f) = \mathbb{E}_{P_x}[f(x)] - r\mathbb{E}_{Q_x}[f(x)] - s\log\mathbb{E}_{Q_x}[e^{\alpha f(x)}] - t\log\mathbb{E}_{Q_x}[e^{\beta f(x)}]$$

$$\Rightarrow \mathbb{E}_{P_x}[\delta] - r\mathbb{E}_{Q}[\delta] - s\log\underbrace{\mathbb{E}_{Q}[1+\alpha\delta]}_{\alpha\mathbb{E}_Q[\delta]} - t\log\underbrace{\mathbb{E}_{Q}[1+\beta\delta]}_{\beta\mathbb{E}_Q[\delta]} = 0 \quad \forall\delta \text{ sufficiently small}$$

$$\Rightarrow \mathbb{E}_{P_x}[\delta] - (\cancel{r+s\alpha+\beta t})^{1}\,\mathbb{E}_{Q_x}[\delta] = 0 \;=\; \mathbb{E}_{P_x}[\delta] - \mathbb{E}_{Q_x}[\delta] = 0 \quad \forall\delta \text{ ''} \quad \text{''}$$

$$\Rightarrow \int (P_x - Q_x)\delta\, dx = 0 \;\Rightarrow\; \boxed{P_x = Q_x}$$

$$\Rightarrow \boxed{V_{\alpha,\beta,r,s,t}(P_x\|Q_x) = 0 \quad \text{iff} \quad P_x = Q_x}$$

(4

$$V_{\alpha,\beta,r,s,t}(P_{xy}\|Q_{xy}) = \sup_{f:\mathcal{X}\mathcal{Y}\to\mathbb{R}}\left\{\mathbb{E}_{P_{xy}}[f(x,y)] - r\mathbb{E}_{Q_{xy}}[f(x,y)] - s\log\mathbb{E}_{Q_{xy}}[e^{\alpha f(x,y)}] - t\log\mathbb{E}_{Q_{xy}}[e^{\beta f(x,y)}]\right\}$$

$$\geq \sup_{\substack{f(x,y)=g(x)+1\\ g:\mathcal{X}\to\mathbb{R}}}\left\{\mathbb{E}_{P_{xy}}[f(x,y)] - r\mathbb{E}_{Q_{xy}}[f(x,y)] - s\log\mathbb{E}_{Q_{xy}}[e^{\alpha f(x,y)}] - t\log\mathbb{E}_{Q_{xy}}[e^{\beta f(x,y)}]\right\}$$

$$= \sup_{g:\mathcal{X}\to\mathbb{R}}\left\{\mathbb{E}_{P_{xy}}[g(x)+1] - r\mathbb{E}_{Q_{xy}}[g(x)+1] - s\log\mathbb{E}_{Q_{xy}}[e^{\alpha(g(x)+1)}] - t\log\mathbb{E}_{Q_{xy}}[e^{\beta(g(x)+1)}]\right\}$$

$$= \sup_{g:\mathcal{X}\to\mathbb{R}}\left\{\mathbb{E}_{P_x}[g(x)] - r\mathbb{E}_{Q_x}[g(x)] + 1 - r - s\log\left(e^{\alpha}\mathbb{E}_{Q_x}[e^{\alpha g(x)}]\right) - t\log\left(e^{\beta}\mathbb{E}_{Q_x}[e^{\beta g(x)}]\right)\right\}$$

$$= \sup_{g:\mathcal{X}\to\mathbb{R}}\left\{\mathbb{E}_{P_x}[g(x)] - r\mathbb{E}_{Q_x}[g(x)] - s\log\mathbb{E}_{Q_x}[e^{\alpha g(x)}] - t\log\mathbb{E}_{Q_x}[e^{\beta g(x)}] + \cancel{1-r-s\alpha-t\beta}^{\,0}\right\} = V_{\alpha,\beta,r,s,t}(P_x\|Q_x)$$

$$\Rightarrow \boxed{V_{\alpha,\beta,r,s,t}(P_{xy}\|Q_{xy}) \geq V_{\alpha,\beta,r,s,t}(P_x\|Q_y)} \quad \Leftarrow \quad \sup_{f:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}} \geq \sup_{\substack{f=g(x)+c\\ g:\mathcal{X}\to\mathbb{R}}}$$

(5

$$V_{\alpha,\beta,r,s,t}(\underbrace{P_x W_{x|y}}_{P_{xy}} \| \underbrace{Q_x W_{y|x}}_{Q_{xy}}) = \sup_{f:\mathcal{X}\mathcal{Y}\to\mathbb{R}}\left\{\mathbb{E}_{P_{xy}}[f(x,y)] - r\mathbb{E}_{Q_{xy}}[f(x,y)] - s\log\mathbb{E}_{Q_{xy}}[e^{\alpha f(x,y)}] - t\log\mathbb{E}_{Q_{xy}}[e^{\beta f(x,y)}]\right\}$$

$$= \sup_{f:\mathcal{X}\mathcal{Y}\to\mathbb{R}}\left\{\mathbb{E}_{P_x}\mathbb{E}_{W_{x|y}}[f(x,y)] - r\mathbb{E}_{Q_x}\mathbb{E}_{W_{y|x}}[f(x,y)] - s\log\mathbb{E}_{Q_x}\mathbb{E}_{W_{y|x}}[e^{\alpha f(x,y)}] - t\log\mathbb{E}_{Q_x}\mathbb{E}_{W_{y|x}}[e^{\beta f(x,y)}]\right\}$$

$$\leq \sup_{f:\mathcal{X}\mathcal{Y}\to\mathbb{R}}\left\{\mathbb{E}_{P_x}\mathbb{E}_{W_{x|y}}[f(x,y)] - r\mathbb{E}_{Q_x}\mathbb{E}_{W_{y|x}}[f(x,y)] - s\log\mathbb{E}_{Q_x}[e^{\alpha\mathbb{E}_{W_{y|x}}f(x,y)}] - t\log\mathbb{E}_{Q_x}[e^{\beta\mathbb{E}_{W_{y|x}}f(x,y)}]\right\} : \mathbb{E}_{W_{y|x}}f(x,y) = \bar{f}(x)$$

$$= \sup_{\substack{f:\mathcal{X}\mathcal{Y}\to\mathbb{R}\\ \Updownarrow\\ \bar{f}:\mathcal{X}\to\mathbb{R}}}\left\{\mathbb{E}_{P_x}[\bar{f}(x)] - r\mathbb{E}_{Q_x}[\bar{f}(x)] - s\log\mathbb{E}_{Q_x}[e^{\alpha\bar{f}(x)}] - t\log\mathbb{E}_{Q_x}[e^{\beta\bar{f}(x)}]\right\} = V_{\alpha,\beta,r,s,t}(P_x\|Q_x)$$

$\Rightarrow$ since only $\bar{f}(x)$ matters anyway

$$\Rightarrow V_{\alpha,\beta,r,s,t}(\overset{P_{xy}}{\widetilde{P_x W_{y|x}}} \| \overset{Q_{xy}}{\widetilde{Q_x W_{y|x}}}) \leq V_{\alpha,\beta,r,s,t}(P_x \| Q_y)$$

Part 4: $V_{\alpha,\beta,r,s,t}(P_{xy} \| Q_{xy}) \geq V_{\alpha,\beta,r,s,t}(P_x \| Q_y)$

$$\Rightarrow \boxed{V_{\alpha,\beta,r,s,t}(P_x W_{y|x} \| Q_x W_{y|x}) = V_{\alpha,\beta,r,s,t}(P_x \| Q_x)}$$

DP: $V_{\alpha,\beta,r,s,t}(P_x \| Q_x) = V_{\alpha,\beta,r,s,t}(P_x W_{y|x} \| Q_x W_{y|x}) = V_{\alpha,\beta,r,s,s}(P_{yx} \| Q_{yx}) \geq V_{\alpha,\beta,r,s,t}(P_y \| Q_y)$

$$\Rightarrow \boxed{V_{\alpha,\beta,r,s,t}(P_x \| Q_x) \geq V_{\alpha,\beta,r,s,t}(P_y \| Q_y)} \text{ for } P_{y|x} = Q_{y|x} = W_{y|x}$$

(6

$$V_{\alpha,\beta,r,s,t}(P_{xy} \| Q_x Q_y) = \sup_{f: \mathcal{X}\mathcal{Y} \to \mathbb{R}} \left\{ \mathbb{E}_{P_{xy}}[f(x,y)] - r\mathbb{E}_{Q_x Q_y}[f(x,y)] - s\log \mathbb{E}_{Q_x Q_y}[e^{\alpha f(x,y)}] - t\log \mathbb{E}_{Q_x Q_y}[e^{\beta f(x,y)}] \right\}$$

$$\geq \sup_{\substack{f = g+h \\ g: \mathcal{X} \to \mathbb{R} \\ h: \mathcal{Y} \to \mathbb{R}}} \left\{ \mathbb{E}_{P_{xy}}[f(x,y)] - r\mathbb{E}_{Q_x Q_y}[f(x,y)] - s\log \mathbb{E}_{Q_x Q_y}[e^{\alpha f(x,y)}] - t\log \mathbb{E}_{Q_x Q_y}[e^{\beta f(x,y)}] \right\} \ldots$$

$$= \sup_{g,h: \mathcal{X}\mathcal{Y} \to \mathbb{R}} \left\{ \mathbb{E}_{P_{xy}}[g(x)+h(y)] - r\mathbb{E}_{Q_x Q_y}[g(x)+h(y)] - s\log \underbrace{\mathbb{E}_{Q_x}\mathbb{E}_{Q_y}[e^{\alpha g(x)} \cdot e^{\alpha h(y)}]}_{\mathbb{E}_{Q_x}[e^{\alpha g(x)}] \cdot \mathbb{E}_{Q_y}[e^{\alpha h(y)}]} - t\log \underbrace{\mathbb{E}_{Q_x}\mathbb{E}_{Q_y}[e^{\beta g(x)} \cdot e^{\beta h(y)}]}_{\mathbb{E}_{Q_x}[e^{\beta g(x)}] \cdot \mathbb{E}_{Q_y}[e^{\beta h(y)}]} \right\}$$

$$= \sup_{g,h: \mathcal{X},\mathcal{Y} \to \mathbb{R}} \left\{ \mathbb{E}_{P_x}[g(x)] + \mathbb{E}_{P_y}[h(y)] - r\mathbb{E}_{Q_x}[g(x)] - r\mathbb{E}_{Q_y}[h(y)] - s\log \mathbb{E}_{Q_x}[e^{\alpha g(x)}] - s\log \mathbb{E}_{Q_y}[e^{\alpha h(y)}] - t\log \mathbb{E}_{Q_x}[e^{\beta g(x)}] - t\log \mathbb{E}_{Q_y}[e^{\beta h(y)}] \right\}$$

$$= \sup_{g: \mathcal{X} \to \mathbb{R}} \left\{ \mathbb{E}_{P_x}[g(x)] - r\mathbb{E}_{Q_x}[g(x)] - s\log \mathbb{E}_{Q_x}[e^{\alpha g(x)}] - t\log \mathbb{E}_{Q_x}[e^{\beta g(x)}] \right\} +$$

$$\sup_{h: \mathcal{Y} \to \mathbb{R}} \left\{ \mathbb{E}_{P_y}[h(y)] - r\mathbb{E}_{Q_y}[h(y)] - s\log \mathbb{E}_{Q_y}[e^{\alpha h(y)}] - t\log \mathbb{E}_{Q_y}[e^{\beta h(y)}] \right\} = V_{\alpha,\beta,r,s,t}(P_x \| Q_x) + V_{\alpha,\beta,r,s,t}(P_y \| Q_y)$$

$$\Rightarrow \boxed{V_{\alpha,\beta,r,s,t}(P_{xy} \| Q_x Q_y) \geq V_{\alpha,\beta,r,s,t}(P_x \| Q_x) + V_{\alpha,\beta,r,s,t}(P_y \| Q_y)}$$

(7

$$W_\alpha(P_x \| Q_x) = V_{\alpha,0,1-\frac{1}{\alpha},\frac{1}{\alpha^2},0}(P_x \| Q_x)$$

$$W_\alpha(P_x \| Q_x) = \sup_{f: \mathcal{X} \to \mathbb{R}} \left\{ \mathbb{E}_{P_x}[f(x)] - (1-\tfrac{1}{\alpha})\mathbb{E}_{Q}[f(x)] - \frac{1}{\alpha^2}\log(\mathbb{E}_{Q_x}[e^{\alpha f(x)}]) \right\}$$

using taylor expansion: $e^{\alpha f(x)} \approx 1 + \alpha f(x) + \frac{1}{2}\alpha^2 f(x)^2 + O(\alpha^3)$

$$\Rightarrow \mathbb{E}_Q[e^{\alpha f(x)}] \approx 1 + \alpha \mathbb{E}_Q[f(x)] + \frac{1}{2}\alpha^2 \mathbb{E}_Q[f^2(x)] + O(\alpha^3)$$

$$\log(1+\delta) \approx \delta - \tfrac{1}{2}\delta^2 + O(\delta^3) \Rightarrow \log \mathbb{E}_Q[e^{\alpha f(x)}] \approx \alpha \mathbb{E}_Q[f(x)] + \frac{1}{2}\alpha^2 \mathbb{E}_Q[f^2(x)] - \frac{1}{2}\alpha^2 \mathbb{E}_Q[f(x)]^2 + O(\alpha^3)$$

$\Rightarrow \log \mathbb{E}[e^{\alpha f(x)}] \approx \alpha \mathbb{E}_Q[f(x)] + \frac{1}{2} \mathrm{Var}_Q[f(x)]\, \alpha^2 + O(\alpha^3)$

$$\Rightarrow W_\alpha(P_X \| Q_X) = \sup_{f: \mathcal{X} \to \mathbb{R}} \left\{ \mathbb{E}_P[f(x)] - (1 - \cancel{\tfrac{1}{\alpha}}) \mathbb{E}_Q[f(x)] - \tfrac{1}{\cancel{\alpha}} \cancel{\mathbb{E}_Q}[f(x)] - \tfrac{1}{2} \mathrm{Var}_Q[f(x)] + O(\alpha) \right\}$$

$$= \sup_{f: \mathcal{X} \to \mathbb{R}} \left\{ \mathbb{E}_{P_X}[f(x)] - \mathbb{E}_Q[f(x)] - \tfrac{1}{2} \mathrm{Var}_Q[f(x)] + O(\alpha) \right\}$$

$$\Rightarrow \lim_{\alpha \to 0} W_\alpha(P_X \| Q_X) = \sup_{f: \mathcal{X} \to \mathbb{R}} \left\{ \mathbb{E}_{P_X}[f(x)] - \mathbb{E}_Q[f(x)] - \tfrac{1}{2} \mathrm{Var}_Q[f(x)] \right\}$$

$$f \to cf \Rightarrow \lim_{\alpha \to 0} W_\alpha(P_X \| Q_X) = \sup_{f: \mathcal{X} \to \mathbb{R}} \sup_{c \in \mathbb{R}} \left\{ c\left(\mathbb{E}_{P_X}[f(x)] - \mathbb{E}_{Q_X}[f(x)]\right) - \tfrac{c^2}{2} \mathrm{Var}_Q[f(x)] \right\}$$

$$\Rightarrow \frac{\partial}{\partial c} = 0 \to c_{opt} = \frac{\mathbb{E}_{P_X}[f(x)] - \mathbb{E}_{Q_X}[f(x)]}{\mathrm{Var}_{Q_X}[f(x)]} \rightsquigarrow \lim_{\alpha \to 0} W_\alpha(P_X \| Q_X) = \frac{1}{2} \sup_{f: \mathcal{X} \to \mathbb{R}} \frac{\left(\mathbb{E}_{P_X}[f(x)] - \mathbb{E}_{Q_X}[f(x)]\right)^2}{\mathrm{Var}_{Q_X}[f(x)]}$$

$$\rightsquigarrow \boxed{\lim_{\alpha \to 0} W_\alpha(P_X \| Q_X) = \frac{1}{2} \chi^2(P_X, Q_X)}$$

$$\forall \alpha, \beta, r, s, t : r + \alpha s + \beta t = 1 : \quad V_{\alpha,\beta,r,s,t}(P_{XY} \| Q_X Q_Y) \geq V_{\alpha,\beta,r,s,t}(P_X \| Q_X) + V_{\alpha,\beta,r,s,t}(P_Y \| Q_Y)$$

$$,\ \alpha \times \frac{1}{\alpha^2} + (1 - \frac{1}{\alpha}) = 1 \rightsquigarrow V_{\alpha, 1, 1-\frac{1}{\alpha}, \frac{1}{\alpha^2}} \text{ is valid} \Rightarrow \lim_{\alpha \to 0} V_{\alpha, 0, 1-\frac{1}{\alpha}, \frac{1}{\alpha^2}, 0} = \lim_{\alpha \to 0} W_\alpha = \frac{1}{2} \chi^2$$

also works $\Rightarrow \boxed{\cancel{\tfrac{1}{2}} \chi^2(P_{XY} \| Q_X Q_Y) \geq \cancel{\tfrac{1}{2}} \chi^2(P_X \| Q_X) + \cancel{\tfrac{1}{2}} \chi^2(P_Y \| Q_Y)}$ $\rightsquigarrow$ holds.