تمرین سری ۲ نظریه اطلاعات، آمار و یادگیری ITSL   محمدپارسا دینی 400101204

(I) $P \ll Q$; $Z_i = \log\left(\frac{P(X_i)}{Q(X_i)}\right)$; $W_i = \log\left(\frac{P(Y_i)}{Q(Y_i)}\right)$; $X_i \overset{iid}{\sim} P$, $Y_i \overset{iid}{\sim} Q$

$$\mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\right] \underset{\lambda>0}{=} \mathbb{P}\left[e^{\lambda\sum_{i=1}^{n}(W_i-Z_i)} \geq e^{\lambda n t}\right] \underset{\text{مارکوف}}{\leq} e^{-\lambda n t}\,\mathbb{E}_{X,Y^n}\left[e^{\lambda\sum_{i=1}^{n}(W_i-Z_i)}\right]$$

حال با استفاده از iid بودن X_i و Y_i ها و همچنین مستقل بودن X و Y داریم:

$$\mathbb{E}_{X,Y^n}\left[e^{\lambda\sum_{i=1}^{n}(W_i-Z_i)}\right] = \prod_{i=1}^{n}\mathbb{E}_{X_i,Y_i}\left[e^{\lambda(W_i-Z_i)}\right] \underset{\text{(iid بودن)}}{=} \left[\mathbb{E}_{X,Y}\, e^{\lambda(W-Z)}\right]^n$$

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

در نتیجه اگر از R.H.S روی λ inf بگیریم داریم:

$$\Rightarrow \mathbb{P}\left[\sum_{i=1}^{n}(W_i-Z_i)\geq nt\right] \leq \inf_{\lambda>0} \exp\left(n\left(-\lambda t + \log\left(\mathbb{E}_X e^{-\lambda Z}\right) + \log\left(\mathbb{E}_Y e^{\lambda W}\right)\right)\right) = \dots$$

$$= \sup_{\lambda>0} \exp\left(-n\left(\lambda t - \log\left(\mathbb{E}_X e^{-\lambda Z}\right) - \log\left(\mathbb{E}_Y e^{\lambda W}\right)\right)\right) = \exp\Big(-n \cdot \underbrace{\sup_{\lambda>0}\left\{\lambda t - \log\left(\mathbb{E}_X e^{-\lambda Z}\right) - \log\left(\mathbb{E}_Y e^{\lambda W}\right)\right\}}_{F(t)}\Big)$$

$$\mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i)\geq nt\right] \leq \exp(-nF(t)) \quad . \text{ Q.E.D. } \blacksquare$$

[illegible] ←

(II) for $t=0 \longrightarrow$

$$F(0) = \sup_{\lambda>0}\left\{-\psi_P(-\lambda) - \psi_Q(\lambda)\right\} = \sup_{\lambda>0}\left\{-\log\left(\mathbb{E}_X e^{-\lambda Z}\right) - \log\left(\mathbb{E}_Y\left(e^{\lambda W}\right)\right)\right\}$$

$$= -\inf_{\lambda>0}\left\{\log\left(\int_{-\infty}^{\infty}\underbrace{P(x)\left(\frac{Q(x)}{P(x)}\right)^{\lambda}}_{P(x)^{1-\lambda}Q(x)^{\lambda}}dx\right) + \log\left(\int_{-\infty}^{\infty}\underbrace{Q(y)\left(\frac{P(y)}{Q(y)}\right)^{\lambda}}_{Q(y)^{1-\lambda}P(y)^{\lambda}}dy\right)\right\}$$

$$= -\inf_{\lambda>0}\left\{\log\left(\int_{-\infty}^{\infty}P(x)^{1-\lambda}Q(x)^{\lambda}dx \cdot \int_{-\infty}^{\infty}P(y)^{\lambda}Q(y)^{1-\lambda}dy\right)\right\}$$

حال طبق لم Cauchy Shwartz داریم:

$$= \sup_{\lambda>0} -\log\left(\int_{-\infty}^{\infty}\sqrt{P(x)^{1-\lambda+\lambda}Q(x)^{\lambda+1-\lambda}}\,dx\right)^2 \underset{\text{به } \lambda \text{ بستگی ندارد}}{=} -2\log\left(\int_{-\infty}^{\infty}\sqrt{P(x)Q(x)}\,dx\right)$$

از طرف دیگر داریم:

$$F(0) = \sup_{\lambda>0}\left\{-2\log\left(\int_{-\infty}^{\infty}\sqrt{P(x)Q(x)}\,dx\right)\right\} = -2\log\left(\mathbb{E}_{Y\sim Q}\left[\sqrt{\frac{P(Y)}{Q(Y)}}\right]\right)$$

حال ثابت شد که

$$f(0) = \alpha = -2\log B(P,Q) = -2\log\left(\mathbb{E}_{Y\sim Q}\sqrt{\frac{P(Y)}{Q(Y)}}\right)$$

از طرف دیگر می‌دانیم که حالت تساوی برای نامساوی کوشی دقیقاً است که نسبت متغیر - متغیر ثابت باشد:

$$\forall x: \quad C = \frac{P(x)^{1-\lambda} Q(x)^{\lambda}}{P(x)^{\lambda} Q(x)^{1-\lambda}} = P(x)^{1-2\lambda} Q(x)^{1-2\lambda} \longrightarrow 1-2\lambda=0$$
$\exists c$

اما عبارت فوق نمی‌تواند برای همه x ها برقرار باشد چون P و Q از هم مستقل‌اند و دلیلی هم ندارند.
در نتیجه توان $P(x)$ و $Q(x)$ باید صفر باشد که نتیجه می‌دهد $1-2\lambda=0 \rightarrow \lambda=\frac{1}{2}$ که در نتیجه می‌دهد:

$$\longrightarrow f(0) = \alpha = -2\log B(P,Q) = -\psi_P(-\lambda)-\psi_Q(\lambda)\Big|_{\lambda=\frac{1}{2}} = -\psi_P(-1/2)-\psi_Q(1/2).$$

■ Q.E.D.

III $$f(t) = \sup_{\lambda\geq 0}\{\underbrace{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)}_{f(\lambda,t)}\} \geq f(\lambda_0,t)\Big|_{\lambda_0=\frac{1}{2}}$$

$$\xrightarrow{\forall t} f(t) \geq \frac{t}{2} - \underbrace{\psi_P(-\tfrac{1}{2}) - \psi_Q(1/2)}_{f(0)} = \frac{t}{2} + f(0) \Longrightarrow \forall t: f(t) \geq \frac{t}{2} + f(0)$$

$$\Rightarrow \underset{n\in\mathbb{N}}{\forall t}: -nf(t) \leq -n\left(f(0)+t/2\right) \quad (*)$$

حال با استفاده از نتیجه قبلی و نامساوی نتیجه می‌گیریم:

$$\mathbb{P}\left[\sum_{i=1}^{n}(W_i - Z_i) \geq nt\right] \leq \exp(-nf(t)) \overset{(*)}{\leq} \exp\left[-n\left(\frac{t}{2}+f(0)\right)\right]$$

■ Q.E.D.

سوال 2: ابتدا فرض می‌کنیم که $(\varepsilon_0,\varepsilon_1)$ و (f_0,f_1) دو زوج نرخ قابل دسترس باشند به طوری که: $\exists \pi_{110},\pi_{011} ; P_{110}, P_{011}$:

I

$$\hookrightarrow \begin{cases}\pi_{110} \leq 2^{-n\varepsilon_0} \\ P_{110} \leq 2^{-n}\end{cases} ; \begin{cases}\pi_{011} \leq 2^{-nf_1} \\ \pi_{110} \leq 2^{-nf_0}\end{cases}$$

حال نرخ $(\lambda\varepsilon_0+\bar{\lambda}f_0 , \lambda\varepsilon_1+\bar{\lambda}f_1)$ را در نظر می‌گیریم:

$$\left.\begin{array}{l}\varepsilon^*_\lambda = \lambda\varepsilon_0+\bar{\lambda}\varepsilon_1 \leq \max(\varepsilon_0,f_0) \\ f^*_\lambda = \lambda\varepsilon_1+\bar{\lambda}f_1 \leq \max(\varepsilon_1,f_1) \\ \lambda\in[0,1]\end{array}\right\} \Rightarrow \begin{cases}2^{-n(\lambda\varepsilon_0+\bar{\lambda}f_0)} \geq 2^{-n\max(\varepsilon_0,f_0)} \geq 2^{-n\varepsilon_0}, 2^{-nf_0} \\ 2^{-n(\lambda\varepsilon_1+\bar{\lambda}f_1)} \geq 2^{-n\max(\varepsilon_1,f_1)} \geq 2^{-n\varepsilon_1}, 2^{-nf_1}\end{cases}$$

در نتیجه همان مقادیر π و P نیز در معادلات فوق صدق می‌کنند و در نتیجه چون π ای یافت شد که زوج نرخ $(\varepsilon^*_\lambda,f^*_\lambda)$ برای هر $\lambda\in(0,1)$ به شرطی که زوج نرخ‌های $(\varepsilon_0,\varepsilon_1)$ و (f_0,f_1) در دسترس باشند؛ می‌توان نتیجه گرفت

■ Q.E.D.

که ناحیه زوج نرخ‌های در دسترس، یک ناحیه محدب باشد. البته چون که اپسیلون π_{11} و π_{011} نیز محدب بود و π را طبق تبدیل $\frac{1}{\pi}$ و $\frac{1}{n}$ به ε تبدیل می‌شوند که محدب است، می‌توان نتیجه گرفت که $(\varepsilon_0,\varepsilon_1)$ نیز محدب است.

2

از استدلال 2: می‌دانیم که

(i) (II)

$$\Pi_{1|0} = P\Big(\overbrace{\sum_{k=1}^{n} \log\Big(\frac{Q(X_k)}{P(X_k)}\Big)}^{T(x^n)} \geq \tau = nt\Big) \xrightarrow{\lambda>0} = P\Big(e^{\lambda \sum_{k=1}^{n} \log_2\big(\frac{Q(X_k)}{P(X_k)}\big)} \geq e^{\lambda nt}\Big)$$

$$\to \text{By Markov} \Rightarrow \quad e^{-\lambda nt}\, \mathbb{E}_{X_k\sim P}\Big[e^{\lambda\sum_{k=1}^{n} \log\big(\frac{Q(X_k)}{P(X_k)}\big)}\Big] \geq P\big(T(x^n) \geq nt\big)$$

$$\text{let: } \psi_P(\lambda) = \log\Big(\mathbb{E}_{X\sim P}\, e^{\lambda \log\big(\frac{Q(X)}{P(X)}\big)}\Big)$$

حال اگر inf بگیریم از L.H.S. فشرده‌تر است:

$$P\big(T(x^n) \geq nt\big) \leq \inf_{\lambda>0} e^{-\lambda nt}\, \mathbb{E}_{X\sim P}\Big[e^{\lambda \log_2\big(\frac{Q(X)}{P(X)}\big)}\Big]^n = \exp\Big[-n \cdot \sup_{\lambda>0}\{\lambda t - \psi_P(\lambda)\}\Big]$$

$$\Rightarrow P\big(T(x^n) \geq nt\big) \leq e^{-n\psi_P^*(t)}$$

Q.E.D. ■

$$\text{let } \psi_P^*(t) = \sup_{\lambda>0}\{\lambda t - \psi_P(\lambda)\}$$

ما در بالا از کران مارکف استفاده کردیم که شرط آن این است.

(ii)

$$\Pi_{0|1} = Q\Big(\overbrace{\sum_{k=1}^{n} \log\Big(\frac{Q(X_k)}{P(X_k)}\Big)}^{T(x^n)} < \tau = nt\Big) = Q\Big(e^{-\lambda\sum_{k=1}^{n} \log_2\big(\frac{Q(X_k)}{P(X_k)}\big)} \geq e^{\lambda nt}\Big)$$

$$\to \text{By Markov} \Rightarrow \quad e^{-\lambda nt}\, \mathbb{E}_{X_k\sim Q}\Big[e^{-\lambda\sum_{k=1}^{n} \log\big(\frac{Q(X_k)}{P(X_k)}\big)}\Big] \geq Q\big(T(x^n) \geq nt\big)$$

$$\text{let } \psi_Q(\lambda) = \log\Big(\mathbb{E}_{X\sim P}\Big[e^{\lambda \log\big(\frac{P(X)}{Q(X)}\big)}\Big]\Big)$$

حال اگر inf بگیریم از L.H.S. فشرده‌تر است

$$\Rightarrow Q\big(T(x^n) \not\geq nt\big) \leq \inf_{\lambda>0} e^{-\lambda nt}\, \mathbb{E}_{X\sim Q}\Big[e^{+\lambda \log\big(\frac{P(X)}{Q(X)}\big)}\Big]^n = \exp\Big(+n \cdot \sup_{\lambda>0}\{\lambda t - \psi_Q(\lambda)\}\Big)$$

$$\text{let } \psi_Q^*(t) = \sup_{\lambda>0}\{\lambda t - \psi_Q(\lambda)\}$$

$$\Rightarrow Q\big(T(x^n) \not\geq nt\big) \leq e^{-n\psi_Q^*(t)}$$

Q.E.D. ■

در نهایت اثبات کردیم که:

$$\begin{cases} \Pi_{1|0}^{(n)} \leq e^{-n\psi_P^*(t)} \\ \Pi_{0|1}^{(n)} \leq e^{-n\psi_Q^*(t)} \end{cases} \qquad \begin{aligned} &\forall\, t \geq \mathbb{E}_{X\sim P}\Big[\log\frac{P(X)}{Q(X)}\Big] = -D_{KL}(P\|Q) \\ &\forall\, t \leq \mathbb{E}_{X\sim Q}\Big[\log\frac{P(X)}{Q(X)}\Big] = D_{KL}(Q\|P) \end{aligned}$$

(III)

از قسمت 2C:

$$\rightarrow E_0(t) = \frac{1}{n} \sup_{\pi_{1|0}} \log \pi_{1|0}^{(n)} = \frac{1}{n} \log \left[\exp(-n\psi_P^*(t)) \right] = \psi_P^*(t)$$

$$\rightarrow E_1(t) = \frac{1}{n} \sup_{\pi_{0|1}} \log \pi_{0|1}^{(n)} = \frac{1}{n} \log \left[\exp(-n\psi_Q^*(t)) \right] = \psi_Q^*(t) \quad \underline{(**)}$$

حال ψ_Q^* را برحسب ψ_P^* می‌نویسیم:

$$\psi_Q(\lambda) = \log\left(\mathbb{E}_{x\sim Q}\left[e^{\lambda T(x)} \right] \right) = \log\left(\sum_x e^{\lambda \log \frac{Q(x)}{P(x)}} \cdot Q(x) \right) = \log\left(\sum_x e^{\lambda \log \frac{Q(x)}{P(x)}} Q(x) \cdot \underbrace{\frac{P(x)}{Q(x)} e^{\log \frac{Q(x)}{P(x)}}}_{1} \right)$$

$$T(x) = \log \frac{Q(x)}{P(x)}$$

$\rightarrow$ change of measure:

$$\rightarrow \psi_Q(\lambda) = \log\left(\sum_x e^{(\lambda+1)\log \frac{Q(x)}{P(x)}} P(x) \right) = \mathbb{E}_{x\sim P}\left[e^{(\lambda+1)T(x)} \right] = \psi_P(\lambda+1) \overset{\underline{(*)}}{=} \psi_Q(\lambda)$$

$$\Rightarrow \psi_Q^*(\lambda) = \sup_{\lambda\geq 0} \{ \lambda t - \psi_Q(\lambda) \} \overset{(*)}{=} \sup_{\lambda \geq 0} \{ \lambda t - \psi_P(\lambda+1) \} = \underbrace{\sup_{\lambda\geq 0} \{ (\lambda+1)t - \psi_P(\lambda+1) \}}_{\psi_P^*(t)} - t$$

$$\Rightarrow \psi_Q^*(\lambda) = \psi_P^*(t) - t$$

حال از $\underline{(*)}$ ، $(**)$ نتیجه می‌گیریم:

$$\begin{cases} E_1(t) = \psi_Q^*(t) = \psi_P^*(t) - t \\ E_0(t) = \psi_P^*(t) \end{cases} \quad : \forall -D_{KL}(P\|Q) \leq t \leq D_{KL}(Q\|P)$$

$\underline{Q.E.D.}$ ∎

IV

ادامه سوال ۲: در اینجا می دانیم که

$$\begin{cases} \varepsilon_0 = \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\pi_{1|0}^{(n)}} \le \inf_{Q:\, \mathbb{E}_Q(X) > \gamma} \left\{ D_{KL}(Q\|P) \right\} \\ \varepsilon_1 = \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\pi_{0|1}^{(n)}} \le \inf_{Q:\, \mathbb{E}_Q(X) \le \gamma} \left\{ D_{KL}(Q\|P) \right\} \end{cases}$$

حال می خواهیم روی $\pi_{1|0}$ و $\pi_{0|1}$ کران بالا زدیم و کران converse را اثبات کنیم و بگوییم که اندازه های $\varepsilon_0, \varepsilon_1$ بیشتر بهتر نمی شوند. می دانیم که $\log \frac{1}{\pi_{1|0}^{(n)}}$ را می توان به صورت زیر نوشت:

$$\log\left(\frac{1}{P(X\in\mathcal{E})}\right) = D_{KL}\left(P_{X|X\in\mathcal{E}} \| P_X\right) \qquad \text{let } \hat{P} = P_{X^n \mid \sum X_i > n\gamma}$$

$\xrightarrow{\quad}$ X_i's are iid, KL properties

$$\hookrightarrow \log\left[\frac{1}{\pi_{1|0}^{(n)}}\right] = D_{KL}\left(\hat{P}_{X^n} \| P_{X^n}\right) \ge \inf_{Q_{X^n}:\, \mathbb{E}_Q(\sum X_i) > n\gamma} \left\{ D(Q_{X^n} \| P_{X^n}) \right\} \ge \sum_{j=1}^{n} D(Q_{X_j}\|P) = n\, D(\bar{Q}\|P)$$

Q.E.D

$$\Rightarrow \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\pi_{1|0}} = \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\pi_{1|0}^{(n)}} \ge \inf_{Q:\, \mathbb{E}_Q(X) > \gamma} \cancel{\frac{n}{n}} D_{KL}(Q\|P) = \boxed{\psi_P^*(t) = \varepsilon_0(t)}$$

به طریق مشابه برای $\varepsilon_1(t)$ به دست می آید که:

$$\log \frac{1}{P(X\notin\mathcal{E})} = D_{KL}\left(P_{X|X\notin\mathcal{E}}\right)$$

$$\hookrightarrow \log\left[\frac{1}{\pi_{0|1}^{(n)}}\right] = D_{KL}\left(\hat{P}_{X^n} \| P_{X^n}\right) \ge \inf_{Q_{X^n}:\, \mathbb{E}_Q(\sum X_i) < n\gamma} \left\{ D(Q_{X^n}\|P_{X^n}) \right\} \ge \sum_{j=1}^{n} D(Q_{X_j}\|P) = n\, D(\bar{Q}\|P)$$

$$\Rightarrow \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\pi_{0|1}^{(n)}} = \lim \frac{1}{n} \log \frac{1}{\pi_{0|1}^{(n)}} \ge \inf_{Q:\, \mathbb{E}_Q(X) \le \gamma} \left\{ D_{KL}(Q\|P) \right\} = \psi_Q^*(t) = \boxed{\psi_P^*(t) - t}$$

Q.E.D.

در نتیجه بهینه کانفرس نیز اثبات شد که منحنی $(\varepsilon_0, \varepsilon_1)$ همان منحنی ای از قبل Converse است.

15/

چون می‌بینیم نتیجه صورت سوال بیت مثبت Achievability را نیز اثبات می‌کنیم:
اتکا با صورت سوال این کمک است
(ولی من کمک نکرده لیز نوشتم)

$$\pi_{1|0} = Q\left[\sum_{k=1}^{n} X_i > n\gamma\right] = 1 - o(1)$$

, X_i : iid

$$\xrightarrow{DPI} D_{KL}(P_{X_1 - X_n} \| Q_{X_1 - X_n}) \geq D_{KL}(P_Z \| Q_Z)$$

$$\begin{cases} Z \sim Ber(\varepsilon_1) \\ Q(Z=0) = \pi_1 = 1-\varepsilon \\ P(Z=0) = \pi_0 \simeq e^{-n\varepsilon} \end{cases}$$

$$D_{KL}(P_Z \| Q_Z) = P(Z=0) \log\left(\frac{P(Z=0)}{Q(Z=0)}\right) + P(Z=1) \log\left(\frac{P(Z=1)}{Q(Z=1)}\right)$$

$$= \left[(1-\varepsilon)\log(1-\varepsilon) + \varepsilon \log \varepsilon\right] + \varepsilon \log \frac{1}{1-e^{-n\varepsilon}} + (1-\varepsilon) n\varepsilon$$

$$\approx -H_b(\varepsilon) + (1-\varepsilon) n E$$

$$E_1 = \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\pi_{1|0}} \leq \lim_{n\to\infty} \left(\frac{D(Q_Z \| P_Z) + o(\frac{1}{n})}{1-\varepsilon}\right) \overset{DPI}{\leq} \frac{D(Q_X \| P_X)}{1-\varepsilon} \approx D_{KL}(Q_X \| P_X)$$

حال اگر Q را optimize کنیم به جواب می‌رسیم:

$$E_1 = \lim_{n\to\infty} \sup_{Q : E_Q(X) > \gamma} \frac{-1}{\pi_{1|0}} \leq \inf_{Q : E_Q(X) > \gamma} \left\{ D_{KL}(Q_{X^n} \| P_{X^n}) \right\}$$

Q.E.D.

Achivability

ادامه سؤال ۲: ابتدا π_0, π_1 را ثابت می‌گیریم. می‌دانیم که برای $\pi_{1|0}, \pi_{0|1}$ که $\pi_1+\pi_0=1$ داریم

$$P_e = \pi_0 \pi_{1|0} + \pi_1 \pi_{0|1} \geq \min\{\pi_{0|1}, \pi_{1|0}\} \quad ; \quad \varepsilon = \frac{-1}{n}\log(P_e)$$

حال اگر روی t نیز sup بگیریم خواهیم داشت: (معادلاً $\varepsilon_1 \sim \pi_{1|0}$ و $\varepsilon_0 \sim \pi_{1|0}$ تبدیل می‌شوند. چرا که منفی لگاریتم نزولی است)

$$\varepsilon = \sup_t \inf_{\pi_{0|1}, \pi_{1|0}} \{P_e\} = \sup_t \{\min\{\varepsilon_1, \varepsilon_0\}\}$$

مشخص می‌شود که تعریف ما از ε همان تعریف اولی است که در سؤال داده شده بود و داشتیم:

از طرف دیگر در قسمت قبل اثبات کردیم که:

$$-D_{KL}(P\|Q) \leq t \leq D_{KL}(Q\|P)$$

$$\begin{cases} \varepsilon_1(t) = \varphi_P^*(t) \\ \varepsilon_0(t) = \varphi_Q^*(t) = \varphi_P^*(t) - t \end{cases} \Rightarrow \begin{cases} \varepsilon_0 > \varepsilon_1 & \text{if } t>0 \\ \varepsilon_0 < \varepsilon_1 & \text{if } t<0 \end{cases}$$

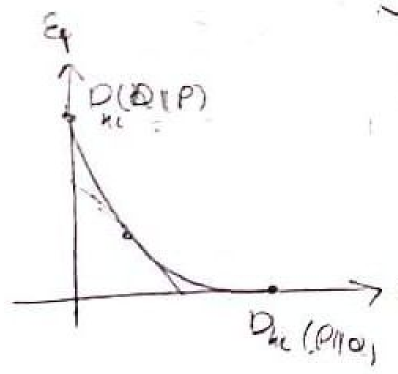
که نتیجه می‌دهد

$$\varepsilon = \sup_t \begin{cases} \varphi_P^*(t) & t>0 \\ \varphi_P^*(t) - t & t<0 \end{cases}$$

ابتدا برای $\varphi_P^*(t)$ که در حالت اول گرفتیم، می‌بینیم که $t=0$ جواب مسأله بیشینه است و به همین ترتیب $\varphi_Q^*(t)$ نیز در $t=0$ رخ می‌دهد (که $\varphi_P^*(0) = \varphi_Q^*(t)\big|_{t=0} = \varphi_P^*(t) - t\big|_{t=0}$) و در نتیجه در هر دو حالت $\varphi_P^*(0)$ جواب بهینه است. اثبات.

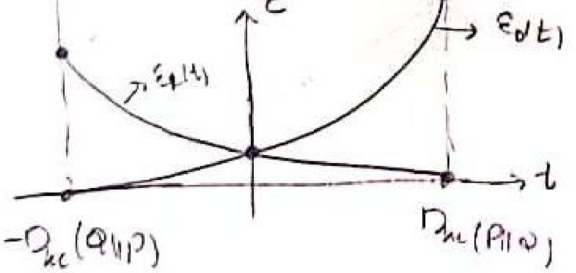
همان طور که می‌دانیم $\begin{cases}\varphi_Q^*(t) \\ \varphi_P^*(t)\end{cases}$ برای $\begin{cases} t<0 \\ t>0 \end{cases}$ صعودی است و برای $\begin{cases} t>0 \\ t<0 \end{cases}$ نزولی است

در نتیجه منحنی $\varepsilon_0(t)$ برحسب t صعودی و برحسب $\varepsilon_1(t)$ نزولی است.



از طرف دیگر بودن در $t=0$ ، $\varphi_Q^*(t)\big|_{t=0} = \varphi_P^*(t)\big|_{t=0}$

که نتیجه می‌گیریم که min فقط در $t=0$ رخ می‌دهد



$$\Rightarrow \varepsilon = \max_t \min_{(\varepsilon_0, \varepsilon_1) \in \xi} \left\{ \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{P_e^{(n)}} \right\} = \boxed{\varphi_P^*(0)}$$

5

Scanned by CamScanner

سوال 3: ابتدا نقاط را به صورت یکنواخت پخش می‌کنیم. در کل 2^n نقطه داریم و احتمال هر یک $\frac{1}{2^n}$ است.
همچنین احتمال اینکه هر مؤلفه بردار تصادفی X، 1 باشد، $\frac{1}{2}$ است. حال می‌خواهیم $f(r_0,n)$ را برحسب احتمال بنویسیم. داریم:

$$f(r_0,n) = 2^n \times \mathbb{P}\left[\|X\|_2 \ge r_0\sqrt{n}\right] = 2^n \times \mathbb{P}\left[\|X\|_2^2 \ge n r_0^2\right] \le \frac{2^n}{n r_0^2}\,\mathbb{E}\left[\|X\|_2^2\right]$$

$$\mathbb{E}\left[\|X\|_2^2\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] = n\,\mathbb{E}\left[X_i^2\right] = \frac{5n}{2} \to f(r_0,n) \le \frac{2^n}{n r_0^2} \times \frac{5n}{2} = \frac{2^n \cdot 5}{2 r_0^2}$$

$$\exp(f(r_0,n)) = \lim_{n\to\infty} \frac{1}{n}\log f(r_0,n) \le \lim_{n\to\infty} \frac{1}{n}\left(n + \log\left(\frac{5}{2r_0^2}\right)\right) = 1 + \lim_{n\to\infty}\frac{1}{n}\log\left(\frac{5}{2r_0^2}\right)$$

حال اگر $r_0^* = o(\sqrt{n})$ باشد، در این صورت $f(r^*,\sqrt{n})$ را حساب کنیم، داریم.

$$f(o(\sqrt{n}),n) \le 1 + \lim_{n\to\infty}\frac{1}{n}\log\frac{5}{2n} = 1 - \lim_{n\to\infty}\frac{\log n}{n} = \boxed{1}$$

از طرف دیگر چون $f(\sqrt{n},n) = 1$ در نتیجه تابع $1+\frac{1}{n}\log\left(\frac{5}{2r_0^2}\right)$ برحسب r_0 اکیداً نزولی است
برای $r_0 < r^*$: $\exp\left[f(r_0,n)\right] = 1$ همچنین برای $r_0 > r^*$ چون که r_0 می‌تواند به اندازه کافی بزرگ شود یعنی $O(r_0) > O(\sqrt{n})$ در این صورت $\frac{-1}{n}\log\left(\frac{5}{2r_0^2}\right)$ منفی‌تر می‌شود و کران بالای exponent از یک کمتر می‌شود.

$$\forall\, r_0 > r^*: \exp\left[f(r_0,n)\right] < 1$$

سوال ۴۰: از قدیم می‌دانیم چیزی که $P_e^{(n)}$ را کمینه می‌کند یک $P_{Z|X}$ یقینی است و به عبارت دیگر باید T threshold

I

کمترین به صورت بهینه P_1 و P_2 را جدا کرد: اثبات:

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

$$\min_{A^n \subseteq \mathcal{X}^n} P_\varepsilon^{(n)} = \pi_1 + \min_{A^{(n)} \subseteq \mathcal{X}^n} \left\{ \pi_2 P_2^{(n)}(A^{(n)}) - \pi_1 P_1^{(n)}(A^{(n)}) \right\} = \pi_1 + \min_{A^n \subseteq \mathcal{X}^n} \left\{ \sum_{x \in \mathcal{X}^n} \pi_2 P_2^{(n)}(x) - \pi_1 P_1^{(n)}(x) \times \mathbb{1}(x \in A^n) \right\}$$

که قبلاً دیدیم طبق ML به $A_*^{(n)} = \{ x \mid \pi_1 P_1^{(n)}(x) \geq \pi_2 P_2^{(n)}(x) \}$ تبدیل می‌شود:

$$\min_{A^n \subseteq \mathcal{X}^n} P_\varepsilon^{(n)} = \pi_1 - \pi_1 P_1(A_*^{(n)}) + \pi_2 P_2(A_*^{(n)}) = \pi_1 P_1^{(n)}\left[\frac{P_1(x^n)}{P_2(x)} < \frac{\pi_2}{\pi_1} \right] + \pi_2 P_2^{(n)}\left[\frac{P_1(x)}{P_2(x)} > \frac{\pi_2}{\pi_1} \right]$$

از بخش 5,4 سوال 2 دیدیم که

$$\psi_Q^*(0) = \psi_P^*(0) = \min_{P_{Z|X}} P_\varepsilon$$

$$= -\inf_\lambda \psi_{P_1}(\lambda) = -\inf_\lambda \psi_{P_2}(\lambda)$$

$$\begin{cases} \psi_{P_1}^*(0) = \sup_\lambda \{ -\psi_{P_1}(\lambda) \} = -\inf_{\lambda \in \mathbb{R}} \psi_{P_1}(\lambda) \\ \psi_{P_2}^*(0) = -\inf_{\lambda \in \mathbb{R}} \psi_{P_2}(\lambda) \end{cases}$$

حال سعی می‌کنیم عبارات فوق را محاسبه کنیم؛ داریم

$$\begin{cases} \sup_\lambda -\psi_P(\lambda) = \sup_\lambda -\log \mathbb{E}_P[e^{\lambda T}] = -\log\left[\sup_\lambda \mathbb{E}_P[e^{\lambda T}] \right] \\ T = \log \frac{dQ}{dP} \end{cases}$$

اما چون که روی سوپ است، چرا که مقدار λ مستقل است، تابع نیز مع فرض دلخواه مقدار تابع مستقل است. پس از مشتق کمک می‌گیریم:

$$\frac{d}{d\lambda} \psi_P(\lambda) = \frac{d}{d\lambda} \mathbb{E}_{x \sim P}\left[e^{\lambda \log \frac{Q(x)}{P(x)}} \right] = \left[\sum_x P(x) e^{\lambda \log \frac{Q(x)}{P(x)}} \left(\log \frac{Q(x)}{P(x)} \right) \right]$$

$$= \sum_x P(x) \cdot \log\left(\frac{Q(x)}{P(x)} \right) \cdot \left(\frac{Q(x)}{P(x)} \right)^\lambda = 0$$

طرفین تقسیم بر $\sum_{y \in \mathcal{X}} P_1^\lambda(y) P_2^{1-\lambda}(y)$:

$$\text{let } P_\lambda(x) = \frac{P_1^\lambda(x) P_2^{1-\lambda}(x)}{\sum_{y \in \mathcal{X}} P_1^\lambda(y) P_2^{1-\lambda}(y)}$$

$$\Updownarrow$$

$$\sum_x \frac{P(x)^{1-\lambda} Q(x)^\lambda \left(\log \frac{Q(x)}{P(x)} \right)}{\sum_{y \in \mathcal{X}} P_1^\lambda(y) P_2^{1-\lambda}(y)} = 0$$

$$\Updownarrow$$

$$\mathbb{E}_{x \sim P_\lambda}[T] = 0$$

حال می‌توانیم عبارت فوق را به صورت $\mathbb{E}_{x \sim P_\lambda}\left[\log \frac{dP_\lambda}{dP} \cdot \frac{dQ}{dP} \right] = 0$ نوشته و نتیجه بگیریم: $D_{KL}(P_\lambda \| Q) = D_{KL}(P_\lambda \| P)$

حال اگر دوبار با P_1 و P_2 تکرار کنیم، داریم و چون طبق مسئله بهینه‌سازی فوق داریم که:

$$\exists \quad \lambda^* = \arg\max_\lambda \{ \psi_{P_1}^*(\lambda) \} = \arg\max_\lambda \{ \psi_{P_2}^*(\lambda) \}$$

$$O^* = \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{P_e^{(n)}} = \sup_\lambda \{ \psi_{P_1}^*(0) \} = \sup_\lambda \{ \psi_{P_2}^*(0) \} = D_{KL}(P_\lambda \| P_1) = D_{KL}(P_\lambda \| P_2)$$

$\Longrightarrow$ Q.E.D.

ادامه سوال ۴ :

II

$$C(P_1,P_2) = +\min_{0\le\lambda\le1} \underbrace{-\log\left[\mathbb{E}_{X\sim P_2}\left[\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\right]\right]}_{C(\lambda)}$$

می‌دانیم که برای $0\le\lambda\le1$ ، $P_1^{\lambda}P_2^{1-\lambda}$ محدب است. در نتیجه چون $-\log$ نیز محدب است و صعودی که نیز نه است، نتیجه می‌گیریم که $C(P_1,P_2)$ نسبت به λ محدب است و از مشتق‌گیری داریم: $(*)$

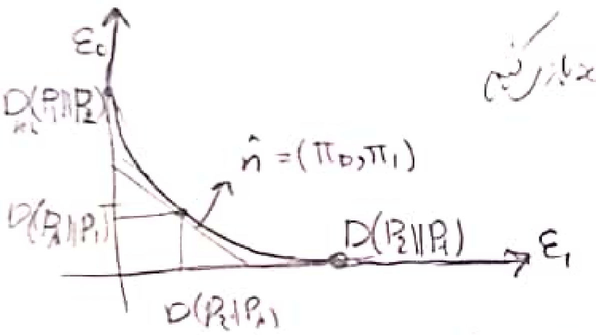
$$\frac{\partial C(\lambda)}{\lambda} = -\frac{\mathbb{E}_{x\sim P_2}\left[\log\left(\frac{P_1(x)}{P_2(x)}\right)\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda}\right]}{\mathbb{E}_{x\sim P_2}\left[\frac{P_1^{\lambda}(x)}{P_2^{\lambda}(x)}\right]} = 0 \Rightarrow \sum_x P_1^{\lambda}(x)P_2^{1-\lambda}(x)\log\left(\frac{P_1(x)}{P_2(x)}\right)=0$$

$$0=\mathbb{E}_{X\sim P_\lambda}\left[\log\frac{P_1(x)}{P_2(x)}\right]$$

را در کنش اول می‌نویسیم که λ^* این خاصیت را داشته باشد:

$$\left(\frac{P_1(x)}{P_2(x)}\right)^{\lambda} = e^{\lambda\log\frac{P_1(x)}{P_2(x)}} \quad (*)$$

$$\mathbb{E}_{X\sim P_{\lambda^*}}\left[\log\left(\frac{P_1(x)}{P_2(x)}\right)\right]=0$$



حال چون هر یک از P_λ و P_1 یکتاست و با می‌توانیم در $x\in X^n$ بیان کنیم

بنابراین از آن رابطه فوق برقرار است، نتیجه می‌گیریم که $P_\lambda=P_{\lambda^*}$ که

و در نتیجه داریم:

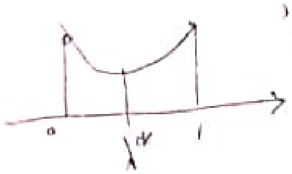
$$\begin{cases} D^* = D_{KL}(P_{\lambda^*}\|P_1) = D_{KL}(P_{\lambda^*}\|P_2) = C(P_1,P_2) \\ C(P_1,P_2) = \min_{\lambda\in[0,1]} -\log\left[\mathbb{E}_{X\sim P_2}\left(\frac{dP_1}{dP_2}\right)^{\lambda}\right] \end{cases}$$

Q.E.D. ∎

دقت می‌کنیم که مسأله محدب بود و اگر در $\lambda=0$ یا $\lambda=1$ نیز نگاه کنیم جواب مسأله کوچکتر بود (زیرا) چرا که:

پس λ^* جواب مسأله بهینه‌سازی بود

$$C(P_1,P_2)\Big|_{\substack{\lambda=0\\ \lambda=1}}=0$$



8

سوال 5:

طبق قضیه Sanov می‌دانیم که:

(i) $\left.\begin{array}{l} X \sim Bin(n,p) \\ X_1, \dots, X_n \overset{iid}{\sim} p \end{array}\right\}$ $\lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \ge \gamma\right)} = \inf_{Q:\, \mathbb{E}_Q(X)\ge\gamma} \left\{ D_{KL}(P\|Q) \right\}$

در نتیجه داریم:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > \gamma\right) \le \exp\left(-n \cdot \inf_{Q:\, \mathbb{E}_Q(X)\ge\gamma} \left\{ D_{KL}(P\|Q) \right\}\right)$$

در نتیجه اگر $X_i \sim Ber(p)$ باشند در این صورت $Y = \sum_{i=1}^{n} X_i \sim Bin(n,p)$ است.

که نتیجه می‌دهد اگر به $\mathbb{P}[\sum_{i=1}^{} X_i \ge k]$ نگاه کنیم باید به جای $\gamma = \frac{k}{n}$ بگذاریم چراکه

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \ge k\right] = \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i > \frac{k}{n}\right] \le \exp\left(-n \cdot \inf_{\mathbb{E}_Q(X)\ge\frac{k}{n}} \left\{ D_{KL}(Q,P) \right\}\right)$$

$\left\{\begin{array}{l} Q \sim Ber(q) \\ P \sim Ber(p) \end{array}\right.$ و در درس دیدیم که این مقدار inf به ازای $\mathbb{E}_Q(X) = \frac{k}{n}$ رخ می‌دهد

$$\hookrightarrow D_{KL}(P\|Q) = d(p\|q) \longrightarrow \inf_{1\ge q\ge\delta\ge p\ge 0} \underbrace{q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}}_{f(q)}$$

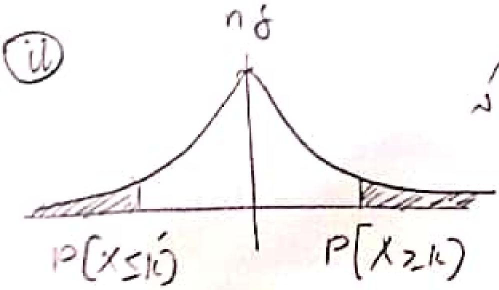
$$\frac{\partial}{\partial q} f(p,q) = 1 + \log q - \log p + \log(1-p) - 1 - \log(1-q) = 0 \longrightarrow \text{either } \frac{q}{1-q} = \frac{p}{1-p} \to q = p$$

$$\hookrightarrow q = 1 \text{ یا } q = \delta$$

$q \ne 1$ و یا $q \ne p$ در نتیجه چون به ازای $q = \delta$ رخ می‌دهد. در نتیجه داریم:

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \ge k\right] \le \exp\left(-n \cdot D_{KL}\left(Ber\left(\tfrac{k}{n}\right) \| Ber(p)\right)\right) \qquad \forall k \ge \mathbb{E}Y = np$$

(ii)



حال از منحنی تقارن ابعاد بزرگ و قضیه حد مرکزی استفاده می‌کنیم:

اگر به توزیع Y وقتی که $n \to \infty$ نگاه کنیم، چون نسبت به وسط آن که np است متقارن است

$$\lim_{n\to\infty} \mathbb{P}[X \ge k] = \lim_{n\to\infty} \mathbb{P}[X \le k'] \qquad k' = n\delta + k$$

$$\overset{n\to\infty}{\Longrightarrow} \mathbb{P}[X \ge k] = \mathbb{P}[X' \le k'] \le \exp\left(-n\, d\left(\tfrac{k}{n}\|p\right)\right) = \exp\left(-n\, D_{KL}\left(Ber\left(\tfrac{k'}{n}\right) \| Ber(p)\right)\right)$$

$$\forall k' \le \mathbb{E}Y = np$$

(i) (II)

ادامه سوال ۵: اگر $d(\frac{k}{n}\|p)$ در قسمت قبل را باز کنیم نتیجه می‌رسیم:

$$\mathbb{P}[X \geq k] \leq \exp\left[-n\cdot\left(\frac{k}{n}\log\frac{k}{np} + \frac{n-k}{n}\log\left(\frac{n-k}{np}\right)\right)\right]$$

now since

$$\forall x,y \in [0,1]^2 : x\log\frac{x}{y} \geq x-y \Rightarrow \left.\begin{array}{l} \text{let } k = u\,\mathbb{E}[X] = unp \\ \text{if } k<np \Rightarrow u<1 \end{array}\right\} \frac{k}{n} = up$$

$$d(\tfrac{k}{n}\|p) = d(up\|p) = up\log u + (1-up)\log\left(\frac{1-up}{1-p}\right)$$

now let $\left.\begin{array}{l} x = 1-up \\ y = 1-p \end{array}\right\}$ $d(\frac{k}{n}\|p) \geq up\log(u) + [1-up-(1-p)] = up\log u + p(1-u)$

$$\left.\begin{array}{l} d(\frac{k}{n}\|p) = p[\underbrace{u\log u + (1-u)}_{f(u)}] = pf(u) \\ \mathbb{E}X = np \end{array}\right\} n\,d(\tfrac{k}{n}\|p) \geq npf(u) = f(u)\,\mathbb{E}[X]$$

$$\Rightarrow \mathbb{P}[X\geq k] \leq \exp(-n\,d(\tfrac{k}{n}\|p)) \leq \exp[-n\cdot f(u)\,\mathbb{E}X] \quad (0<u<1)$$

$$\lim_{n\to\infty} \mathbb{P}[X\leq k) \leq \exp(-n\,d(\tfrac{k}{n}\|p)) \leq \exp(-n\cdot f(u)\,\mathbb{E}X)$$

به روش مشابه می‌توانیم نشان دهیم که

اما در حالت دوم چون که $u>1 \quad \left\{\begin{array}{l} k = u\mathbb{E}X = unp \\ \mathbb{E}X \geq np \end{array}\right.$

پس به دست آمد که

$$\left\{\begin{array}{l} \mathbb{P}(X\geq k] \leq \exp(-nf(u)\mathbb{E}X) : 1>u>0 \\ \mathbb{P}[X\leq k] \leq \exp(-nf(u)\mathbb{E}X) : u>1 \end{array}\right.$$

Q.E.D. ■

(III)

$$f(u) = \int_1^u \frac{u-x}{x}\,dx \underset{x<u}{\geq} \int_1^u \frac{u-x}{u}\,dx = \frac{1}{u}\int_1^u (u-x)\,dx = \frac{(u-1)^2}{2u}$$ Q.E.D ■

حال با توجه به $f(u) \leq \frac{(u-1)^2}{2u}$ و قسمت قبل نتیجه می‌گیریم

$$\mathbb{P}(X\geq u\mathbb{E}X) \leq \exp(-f(u)\mathbb{E}X) \leq \exp\left(\frac{(u-1)^2}{2u}\mathbb{E}X\right) \rightarrow u>1 \Rightarrow \hat{u} = \frac{u}{\mathbb{E}X}$$

now let $u = \mathbb{E}X + t = np+t$

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

10

$$\mathbb{E}X = np \longrightarrow \mathbb{P}[X\geq np+t] \leq \exp\left(\frac{-t^2}{2(t+np)}\right)$$ Q.E.D. ■

ادامه سوال ۵:

(IV)

$$D_{KL}(P\|Q) = \sum_x P(x)\log\frac{P(x)}{Q(x)} = 2\sum_x P(x)\log\sqrt{\frac{P(x)}{Q(x)}} = -2\sum_x P(x)\log\sqrt{\frac{Q(x)}{P(x)}}$$

since $(\log u \le u-1)\ \forall u$:$\Rightarrow D_{KL}(P\|Q) \ge 2\sum_x P(x)\left[1-\sqrt{\frac{Q(x)}{P(x)}}\right]$

$u = \frac{Q(x)}{P(x)} \longrightarrow Q \ll P$ فرض

$$\Rightarrow D_{KL}(P\|Q) \ge 2\sum_x \sqrt{P(x)}\left(\sqrt{P(x)}-\sqrt{Q(x)}\right) \ge 2\sum_x \left(\sqrt{P(x)}-\sqrt{Q(x)}\right)\left(\sqrt{P(x)}-\sqrt{Q(x)}\right)^2$$

① $\sqrt{P(x)} \ge \sqrt{P(x)}-\sqrt{Q(x)}$

② $u>0 \Rightarrow 2u \ge u$

$$\Rightarrow D_{KL}(P\|Q) \ge \sum_x\left(\sqrt{P(x)}-\sqrt{Q(x)}\right)^2 \longrightarrow \left.\begin{array}{l}P\sim Ber(p)\\ Q\sim Ber(q)\end{array}\right\} d(p\|q) \ge (\sqrt{p}-\sqrt{q})^2$$

let $q = k/n \Rightarrow d(\frac{k}{n}\|p) \ge (\sqrt{\frac{k}{n}}-p)^2$

$$\begin{cases}\mathbb{P}[X\ge k] \le \exp(-n\, d(\frac{k}{n}\|p))\\ \mathbb{P}[X\le k] \le \exp(-n\, d(\frac{k}{n}\|p))\end{cases}$$

حال از بخش (I) داریم:

$\boxed{\begin{array}{c}\sqrt{k} = t+\mathbb{E}\sqrt{X} = \sqrt{np}+t\\ t>0\ ;\ \sqrt{k}>\sqrt{np}\end{array}}$

حال ثابت میکنیم

(i)

$$\mathbb{P}[\sqrt{X}\ge\sqrt{k}] = \mathbb{P}[\sqrt{X}-\sqrt{np}\ge t] \le \exp(-n\cdot d(\tfrac{k}{n}\|p)) \le \exp(-n(\sqrt{\tfrac{k}{n}}-\sqrt{p})^2)$$

$$\Rightarrow \mathbb{P}[X\ge k] \le \exp\left(-n\left(\frac{\sqrt{np}+t}{\sqrt{n}}-\sqrt{p}\right)^2\right) = \exp\left(\frac{-nt^2}{n}\right) = \boxed{\exp(-t^2)}$$

Q.E.D.

حال حالت دوم را بررسی میکنیم:

(ii) $\mathbb{P}(X\le k) = \mathbb{P}[\sqrt{X}\le\sqrt{k}]$

$\boxed{\begin{array}{c}\sqrt{k} = -t+\mathbb{E}\sqrt{X} = -t\\ \sqrt{k}<\sqrt{np}\\ t>0\end{array}}$

$$\le \mathbb{P}[\sqrt{X}-\sqrt{np}\le -t] \le \exp(-n\cdot d(\tfrac{k}{n}\|p))$$

$$\le \exp(-n(\sqrt{\tfrac{k}{n}}-\sqrt{p})^2) = \exp(\frac{-nt^2}{n}) = \boxed{\exp(-t^2)}$$

Q.E.D.

در نتیجه به دست آمد که:

$$\begin{cases}\mathbb{P}[\sqrt{X}-\sqrt{np}\ge t] \le e^{-t^2}\\ \mathbb{P}[\sqrt{X}-\sqrt{np}\ge -t] \le e^{-t^2}\end{cases} \quad t>0$$

سوال ۵: فرض می‌کنیم که f^*، بیشینه کننده $V_{\alpha,\beta,r,s,t}$ باشد. داریم:

(I) $\mathcal{X}^{\mathbb{R}} \Rightarrow$ (تمام توابع از $\mathcal{X}$ به $\mathbb{R}$)

$$\sup_{f\in\mathcal{X}^{\mathbb{R}}} V_{\alpha,\beta,r,s,t}(P_x\|Q_x) = V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f^*} \geq V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f} \quad : \forall f\in\mathcal{X}^{\mathbb{R}}$$

حال می‌دانیم که در فرم Variational باید داشته باشیم

$$V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f^*} = V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f^*+c}$$

(به عبارت دیگر شیفت عمودی تاثیری نداشته باشد.)

$$V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f^*+c} = \mathbb{E}_{P_x}\left[f^*(x)+c\right] - r\,\mathbb{E}_{Q_x}\left[f^*(x)+c\right] - s\log\left(\mathbb{E}_{Q_x}\left[\exp\left(\alpha f(x)+\alpha c\right)\right]\right) - t\log\left(\mathbb{E}_{Q_x}\left[\exp\left(\beta f(x)+c\beta\right)\right]\right)$$

$$= c + \mathbb{E}_{P_x}\left[f^*(x)\right] - rc - r\,\mathbb{E}_{Q_x}\left[f^*(x)\right] - sc\alpha - s\log\left(\mathbb{E}_{Q_x}\exp(\alpha f(x))\right) - tc\beta - t\log\left(\mathbb{E}_{Q_x}\exp(\beta f(x))\right)$$

$$= V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f^*} + c\,\underbrace{(1-r-\alpha s-\beta t)}$$

حال اگر ترم $C(1-r-\alpha s-\beta t)$ صفر نباشد، در این صورت اگر C را به سمت بی‌نهایت [illegible] و $V_{\alpha,\beta,r,s,t}=\infty$ و در این صورت [illegible] خوش تعریف نیست.

$$V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f^*} = \begin{cases} \infty & : \forall c,\ 1\neq r+\alpha s+\beta t \\ V_{\alpha,\beta,r,s,t}(P_x\|Q_x)\Big|_{f^*+c} & : \forall c,\ r+\alpha s+\beta t=1 \end{cases}$$

طبق مرحله قبل

(II) می‌دانیم که تابع exp اکیداً محدب است در نتیجه $\mathbb{E}\exp(X)\geq\exp(\mathbb{E}X)$، بنابراین داریم:

sup را [illegible] روی تابع دلخواه کار می‌کنیم که محدب باشد:

$$V_{\alpha,\beta,r,s,t}(P_x\|Q_x) = \sup_{f\in\mathcal{X}^{\mathbb{R}}} V_{\alpha,\beta,r,s,t}(P_x\|Q_x)$$

از نامساوی Jensen داریم:

$$\geq \sup_{f\in\mathcal{X}^{\mathbb{R}}}\left\{\mathbb{E}_{P_x}[f(x)] - r\,\mathbb{E}_{Q}[f(x)] - s\log\left(\exp\left(\alpha\,\mathbb{E}_{Q_x} f(x)\right)\right) - t\log\left(\exp\left(\beta\,\mathbb{E}_{P_x} f(x)\right)\right)\right\}$$

$$= \sup_{f\in\mathcal{X}^{\mathbb{R}}}\left\{\mathbb{E}_{P_x}[f(x)] - \mathbb{E}_{Q_x}[f(x)](r+\alpha s+\beta t)\right\} \geq C\left(1-(\alpha s+\beta t+r)\right) = 0$$

از [illegible] سوپریمم [illegible] کمتر [illegible] نمی‌شود $\Leftarrow \sup_{f\in\mathcal{X}^{\mathbb{R}}} \mathcal{H}(P_x, Q_x; f) \geq \mathcal{H}(P_x, Q_x, f(x)=c)$ تابع ثابت

$$\Longrightarrow V_{\alpha,\beta,r,s,t}(P_x\|Q_x) \geq 0$$

Q.E.D

III ادامه سوال 6: در قسمت قبل به دست آمد که $V_{\alpha,\beta,r,s,t}(P_X \| Q_X) \geq 0$ (∗) حال اگر $Q_X = P_X$ داریم:

$$V_{\alpha,\beta,r,s,t} \doteq \sup_{f \in X^{\mathbb{R}}} \left\{ \mathbb{E}_P[f(x)](r+1) - s\log\left(\mathbb{E}_{P_X}\left[e^{\alpha f(x)}\right]\right) - t\log\left(\mathbb{E}_{P_X}\left[e^{\beta f(x)}\right]\right) \right\}$$

$$\xrightarrow{\text{Jensen}} \quad \leq \mathbb{E}_{P_X}[f(x)]\left(1 - r - \alpha s - \beta t\right) = 0 \quad (\ast\ast)$$

Q.E.D ■

$$\xrightarrow{(\ast),(\ast\ast)} \left(V_{\alpha,\beta,r,s,t}(P_X \| P_X) \geq 0\right) \wedge \left(V_{\alpha,\beta,r,s,t}(P_X \| P_X) \leq 0\right) \Rightarrow V_{\alpha,\beta,r,s,t}(P_X \| P_X) = 0$$

IV $f \in (X \times Y)^{\mathbb{R}}$

ابتدا عبارت شرطی را می‌نویسیم عبارات را در امید دو متغیره ۲ به شکل امید شرطی می‌نویسیم
همچنین فرض داریم که $s\alpha + \beta t + r = 1$ برقرار است.

$$V_{\alpha,\beta,r,s,t}(P_{XY} \| Q_{XY}) = \sup_{f \in (X \times Y)^{\mathbb{R}}} \left\{ \mathbb{E}_{P_{XY}}[f(X,Y)] - r\,\mathbb{E}_{Q_{XY}}[f(X,Y)] - s\log\left[\mathbb{E}_{Q_{XY}} e^{\alpha f(X,Y)}\right] - t\log\left[\mathbb{E}_{Q_{XY}} e^{\beta f(X,Y)}\right] \right\}$$

$$= \sup_{f \in (X \times Y)^{\mathbb{R}}} \left\{ \mathbb{E}_{P_X}\left[\mathbb{E}_{P_{Y|X}} f(X,Y)\right] - r\,\mathbb{E}_{Q_X}\left[\mathbb{E}_{Q_{Y|X}} f(X,Y)\right] - s\log\left[\mathbb{E}_{Q_X}\left[\mathbb{E}_{Q_{Y|X}} e^{\alpha f(X,Y)}\right]\right] - t\log\left[\mathbb{E}_{Q_X}\left[\mathbb{E}_{Q_{Y|X}} e^{\beta f(X,Y)}\right]\right] \right\}$$

فضای کوچکتر

$$\geq \sup_{f(x,y) = g(x) + c} \left\{ \mathbb{E}_{P_X}\left[\underbrace{\mathbb{E}_{P_{Y|X}}[g(x)+c]}_{g(x)+c}\right] - r\,\mathbb{E}_{Q_X}\left[\underbrace{\mathbb{E}_{Q_{Y|X}} g(x)+c}_{g(x)+c}\right] - s\log\left[\mathbb{E}_{Q_X}\left[\underbrace{\mathbb{E}_{Q_{Y|X}} e^{\alpha g(x)+\alpha c}}_{e^{\alpha g(x)+\alpha c}}\right]\right] - t\log\left[\mathbb{E}_{Q_X}\left[\underbrace{\mathbb{E}_{Q_{Y|X}} e^{\beta g(x)+\beta c}}_{e^{\beta g(x)+\beta c}}\right]\right] \right\}$$

$$\geq \sup_{g \in X^{\mathbb{R}}} \left\{ \mathbb{E}_{P_X}[g(x)+c] - r\,\mathbb{E}_{Q_X}[g(x)+c] - s\log\left[\mathbb{E}_{Q_X} e^{\alpha g(x)}\right] - s\alpha c - t\log\left[\mathbb{E}_{Q_X} e^{\beta g(x)}\right] - \beta c t \right\}$$

$$= V_{\alpha,\beta,r,s,t}(P_X \| Q_X) + c(\underbrace{1 - r - \alpha c - \beta t}_{\text{صفر}}) = V(P_X \| Q_X)$$

در نهایت ثابت شد که:

$$\boxed{V_{\alpha,\beta,r,s,t}(P_{XY} \| Q_{XY}) \geq V_{\alpha,\beta,r,s,t}(P_X \| Q_X)}$$

که شبیه نکته اثبات‌ها است که اکنون دیدیم.

Q.E.D ■

V

ادامه سوال ۶b: به روشی که در قسمت III استفاده کردیم متوسل می‌شویم.

از قسمت قبل می‌دانیم که

$$(*)\quad V(P_X\|Q_X) \le V(\hat{P}_{XY}\|\hat{Q}_{XY})$$

کافیست ثابت کنیم که به ازای $\hat{P}_{XY}=P_X W_{Y|X}$ و $\hat{Q}_{XY}=Q_X W_{Y|X}$ داریم: $V(P_X\|Q_X) \ge V(\hat{P}_{XY}\|\hat{Q}_{XY})$ تا نتیجه بگیریم $V(P_X\|Q_X) = V(P_X W_{Y|X}\|Q_X W_{Y|X})$.

$$V_{\alpha,\beta,r,s,t}(\hat{P}_{XY}\|\hat{Q}_{XY}) = V_{\alpha,\beta,r,s,t}(P_X W_{Y|X}\|Q_X W_{Y|X}) =$$

$$= \sup_{f\in(X\times Y)^{\mathbb{R}}} \left\{ \mathbb{E}_{P_X}\left[\mathbb{E}_{W_{Y|X}} f(X,Y)\right] - r\,\mathbb{E}_{Q_X}\left[\mathbb{E}_{W_{Y|X}} f(X,Y)\right] - s\log\left(\mathbb{E}_{Q_X}\left[\mathbb{E}_{W_{Y|X}} e^{\alpha f(X,Y)}\right]\right) - t\log\left(\mathbb{E}_{Q_X}\left[\mathbb{E}_{W_{Y|X}} e^{\beta f(X,Y)}\right]\right)\right\}$$

let $\mathbb{E}_{W_{Y|X}}\left[f(X,Y)\right] = g(X)$

$$\Longrightarrow\ \le \sup_{g\in X^{\mathbb{R}}} \Big\{ \underbrace{\mathbb{E}_{P_X}\left[g(X)\right] - r\,\mathbb{E}_{Q_X}\left[g(X)\right] - s\log\left(\mathbb{E}_{Q_X}\left[e^{\alpha g(X)}\right]\right) - t\log\left(\mathbb{E}_{Q_X}\left[e^{\beta g(X)}\right]\right)}_{V_{\alpha,\beta,r,s,t}(P_X\|Q_X)} \Big\}$$

Q.E.D. ■

در نتیجه ثابت شد که

$$V_{\alpha,\beta,r,s,t}(P_X\|Q_X) \ge V_{\alpha,\beta,r,s,t}(P_X W_{Y|X}\|Q_X W_{Y|X}) \overset{(*)}{\Longrightarrow} \boxed{V(P_X\|Q_X) = V(P_X W_{Y|X}\|Q_X W_{Y|X})}$$

c)

حال باید ثابت کنیم که اگر P_X و Q_X را داشته باشیم هر ترکیب محدب از آنها ($P_\lambda = \lambda P + \bar{\lambda} Q$) که $\lambda\in(0,1)$ را نیز داریم.

let $Z \sim \mathrm{Ber}(\lambda)$

$$P_{X|Z} = \begin{cases} P_\lambda & Z=0 \\ P_{\bar\lambda} & Z=1 \end{cases} \qquad Q_{X|Z} = \begin{cases} Q_\lambda & Z=0 \\ Q_{\bar\lambda} & Z=1 \end{cases}$$

$$\Longrightarrow V(P_{XZ}\|Q_{XZ}) \ge \mathbb{E}_{P_Z}\left[V(P_{X|Z}\|Q_{X|Z})\right] = \lambda V(P_\lambda\|Q_\lambda) + \bar\lambda V(P_{\bar\lambda}\|Q_{\bar\lambda})$$

$$\overset{DPI}{\Longrightarrow} V(P_{XZ}\|Q_{XZ}) \ge V(\underbrace{P_X\|Q_X}_{\text{marginal}}) = V(\underbrace{\lambda P_\lambda + \bar\lambda P_{\bar\lambda} \,\|\, \lambda Q_\lambda + \bar\lambda Q_{\bar\lambda}}_{\lambda(P,Q)+\bar\lambda(P,Q)})$$

حال از بالا در می‌یابیم که $V_{\alpha,\beta,r,s,t}(P_X\|Q_X)$ با شرط $r+s\alpha+\beta t=1$ برای زوج توزیع‌های (P,Q) محدب است.

Q.E.D ■

(VI) در اینجا نیز به مانند قسمت قبل سؤال را حل می‌کنیم: داریم:

$$f(x,y) = h(x) + g(y)$$

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

فقط تابع‌های h, g

$$\geq \sup_{\substack{h\in\mathcal{X}^{\mathbb{R}}\\ g\in\mathcal{Y}^{\mathbb{R}}}} \left\{ \mathbb{E}_{P_{XY}}[h(x)+g(y)] - r\left(\mathbb{E}_{Q_XQ_Y} h(x)+g(y)\right) - s\log\left(\mathbb{E}_{Q_XQ_Y}\left[e^{\alpha h(x)+\alpha g(y)}\right]\right) - t\log\left(\mathbb{E}_{Q_XQ_Y} e^{\beta h(x)+\beta g(y)}\right)\right\}$$

$h(x)$ تنها به X و $g(y)$ تنها به Y بستگی دارد

X و Y مستقل از هم

طبق خاصیت بالا داریم

$$= \sup_{g,h\in\mathcal{X}^{\mathbb{R}}} \left\{ \begin{array}{l} \mathbb{E}_{P_X}[h(x)] - r\,\mathbb{E}_{Q_X}[h(x)] - s\log\left(\mathbb{E}_{Q_X} e^{\alpha h(x)}\right) - t\log\left[\mathbb{E}_{Q_X} e^{\beta h(x)}\right] \\ + \mathbb{E}_{P_Y}[g(y)] - r\,\mathbb{E}_{Q_Y}[g(y)] - s\log\left(\mathbb{E}_{Q_Y} e^{\alpha g(y)}\right) - t\log\left[\mathbb{E}_{Q_Y} e^{\beta g(y)}\right] \end{array}\right\}$$

$$= \sup_{g\in\mathcal{Y}^{\mathbb{R}}}(-) + \sup_{h\in\mathcal{X}^{\mathbb{R}}}(-)$$

چون دو فضای جدا و مستقل از هم هستند پس می‌توان sup ها را جدا کرد

$$= V_{\alpha,\beta,r,s,t}(P_X\|Q_X) + V_{\alpha,\beta,r,s,t}(P_Y\|Q_Y)$$

که در نهایت نتیجه گرفتیم:

$$\boxed{V_{\alpha,\beta,r,s,t}(P_{XY}\|Q_XQ_Y) \geq V_{\alpha,\beta,r,s,t}(P_X\|Q_X) + V_{\alpha,\beta,r,s,t}(P_Y\|Q_Y)}$$

که این خاصیت برای D_{KL} نیز برقرار بود.

(VII)

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

$$= \sup_{f\in\mathcal{X}^{\mathbb{R}}} \left\{ \mathbb{E}_{P_X}[f(x)] - \left(1-\frac{1}{\alpha}\right)\mathbb{E}_{Q_X}[f(x)] - \frac{1}{\alpha^2}\log\left(\mathbb{E}_{Q_X} e^{\alpha f(x)}\right)\right\}$$

ابتدا توجه می‌کنیم که

$$\mathbb{E}_{Q_Y}\left[e^{\alpha f(x)}\right] = \mathbb{E}_{Q_Y}\left[\sum_{k=0}^{\infty}\frac{\alpha^k f(x)^k}{k!}\right] = \sum_{k=0}^{\infty}\frac{\alpha^k}{k!}\mathbb{E}_{Q_X}\left[f(x)^k\right] \longrightarrow f(x) \text{ کران‌دار}$$

از آنجا که $\alpha\to 0$ قرار است برود نیاز نیست بیشتر از مرتبه ۳ بسط تیلور را بیاوریم:

$$\mathbb{E}_{Q_X}\left[e^{\alpha f(x)}\right] = 1 + \alpha\,\mathbb{E}_{Q_X}[f(x)] + \frac{\alpha^2}{2}\mathbb{E}_{Q_X}[f(x)^2] + O(\alpha^3) = 1$$

$$\log(1+t) = \sum_{k=1}^{\infty}\frac{t^k(-1)^{k+1}}{k} \longrightarrow \log(1+t) = t - \frac{t^2}{2} + O(t^3)$$

$$\Rightarrow \log\left(\mathbb{E}_{Q_X} e^{\alpha f(x)}\right) = \log\Big(1 + \underbrace{\alpha\,\mathbb{E}_{Q_X}[f(x)] + \frac{\alpha^2}{2}\mathbb{E}_{Q}[f(x)^2]}_{t} + O(\alpha^3)\Big)$$

$$\Rightarrow \log\left(\mathbb{E}_{Q_x} e^{\alpha f(x)}\right) = \left(\alpha \mathbb{E}_{Q_x}[f(x)] + \frac{\alpha^2}{2}\mathbb{E}_{Q_x}[f(x)^2] + O(\alpha^3)\right) - \frac{1}{2}\left(\alpha \mathbb{E}_{Q_x}[f(x)] + \frac{\alpha^2}{2}\mathbb{E}_{Q_x}[f(x)^2] + O(\alpha^3)\right)^2$$

$$O\left(\left[\alpha \mathbb{E}_{Q}[f(x)] + \frac{\alpha^2}{2}\mathbb{E}_{Q_x}[f(x)^2] + O(\alpha^3)\right)\right)^3$$

$$\Rightarrow \log\left(\mathbb{E}_{Q_x} e^{\alpha f(x)}\right) = \alpha \mathbb{E}_{Q_x}[f(x)] + \frac{\alpha^2}{2}\mathbb{E}_{Q_x}[f(x)^2] + O(\alpha^3) + \frac{O(\alpha^6)}{2} - \frac{\alpha^4}{4}\mathbb{E}_{Q_x}[f(x)^2]^2 + \frac{\alpha^2}{2}\mathbb{E}_{Q_x}[f(x)]^2 + O(\alpha^9)$$

$$\Rightarrow \log\left(\mathbb{E}_{Q_x} e^{\alpha f(x)}\right) \approx \alpha \mathbb{E}_{Q_x}[f(x)] + \frac{\alpha^2}{2}\left(\mathbb{E}_{Q_x}[f(x)^2] - \left(\mathbb{E}_{Q_x}(f(x))\right)^2\right) + O(\alpha^3)$$

$$= \alpha \mathbb{E}_{Q_x}[f(x)] + \frac{\alpha^2}{2}\mathrm{Var}_{Q_x}(f(x)) + O(\alpha^3)$$

$$\lim_{\alpha\to 0} W_\alpha(P_x \| Q_x) = \sup_{\substack{f\in X^{\mathbb{R}} \\ \alpha\to 0}} \left\{ \mathbb{E}_{P_x}[f(x)] - \left(1-\frac{1}{\alpha}\right)\mathbb{E}_{Q}(f(x)) - \frac{1}{\alpha^2}\left(\alpha \mathbb{E}_{Q}(f(x)) + \frac{\alpha^2}{2}\mathrm{Var}_{Q_x}(f(x))\right)\right\}$$

$$= \lim_{\alpha\to 0} \sup_{f\in X^{\mathbb{R}}} \left\{ \mathbb{E}_{P_x}[f(x)] - \mathbb{E}_{Q_x}[f(x)] - \frac{1}{2}\mathrm{Var}_{Q_x}(f(x)) + \cancel{O(\alpha^3)} \right\}$$

حال چون می‌دانیم که دسته توابع ما $cf(x)$ نیز جزو آن قرار می‌گیرند، پس می‌توانیم از آن به جای f در ابتدا
به جای f ، cf گذاشته و دو عبارت را مساوی هم گذاشته و در نهایت c را پیدا می‌کنیم:

$$\mathbb{E}_{P_x}(f(x)\cdot c) - \mathbb{E}_{Q_x}(c\cdot f(x)) - \frac{1}{2}\mathrm{Var}_{Q_x}(c\cdot f(x)) = \overbrace{c\,\mathbb{E}_{P_x}(f(x)) - c\,\mathbb{E}_{Q_x}(f(x)) - \frac{c^2}{2}\mathrm{Var}_{Q_x}(f(x))}^{g(c)}$$

$$\frac{\partial}{\partial c} g(c) = \mathbb{E}_{P_x}(f(x)) - \mathbb{E}_{Q_x}(f(x)) - c\,\mathrm{Var}_{Q_x}[f(x)] = 0$$

$$\hookrightarrow c^* = \frac{\mathbb{E}_P(f(x)) - \mathbb{E}_{Q_x}(f(x))}{\mathrm{Var}_{Q_x}[f(x)]} \longrightarrow$$

همان چیزی که سر کلاس ثابت کردیم!

$$\Rightarrow \lim_{\alpha\to 0} W_\alpha(P_x \| Q_x) = \sup_{f\in X^{\mathbb{R}}} \left\{ \mathbb{E}_{P_x}[c^* f(x)] - \mathbb{E}_{Q_x}[c^* f(x)] - \frac{1}{2}\mathrm{Var}_{Q_x}[c^* f(x)] \right\}$$

$$\lim_{\alpha\to 0} W_\alpha(P_X\|Q_X) = \sup_{f\in\mathcal{X}^{\mathbb{R}}}\left\{ c^*\left(\mathbb{E}_{P_X}(f(X)) - \mathbb{E}_{Q_X}(f(X))\right) + \frac{c^{*2}}{2}\mathrm{Var}_{Q_X}(f(X))\right\}$$

$$= \sup_{f\in\mathcal{X}^{\mathbb{R}}}\left\{ \frac{\left(\mathbb{E}_{P_X}(f(X)) - \mathbb{E}_{Q}(f(X))\right)^2}{\mathrm{Var}_{Q_X}(f(X))} + \frac{1}{2}\,\frac{\left(\mathbb{E}_{P_X}(f(X)) - \mathbb{E}_{Q_X}(f(X))\right)^2}{\mathrm{Var}_{Q_X}(f(X))}\right\}$$

$$= \frac{1}{2}\sup_{f\in\mathcal{X}^{\mathbb{R}}}\left\{ \frac{\left(\mathbb{E}_{P_X}[f(X)] - \mathbb{E}_{Q_X}[f(X)]\right)^2}{\mathrm{Var}_{Q_X}(f(X))}\right\} = \boxed{\frac{1}{2}\chi^2(P_X\|Q_X)}$$

Q.E.D. ■

که فرمولی است که برای χ^2 به دست آوردیم

VIII

طبق چیزی که در 5 ثابت کردیم: اگر $r+\alpha s+\beta t=1$ آنگاه

$$V_{\alpha,\beta,r,s,t}(P_{XY}\|Q_XQ_Y) \geq V_{\alpha,\beta,r,s,t}(P_X\|Q_X) + V_{\alpha,\beta,r,s,t}(P_Y\|Q_Y) \quad (*)$$

حال اگر $\alpha=\alpha$, $\beta=0$, $r=1-\frac{1}{\alpha}$, $s=\frac{1}{\alpha^2}$, $t=0$ را در نظر بگیریم، طبق چیزی که در بالا به دست آوردیم:

$$\left\{\begin{aligned}
\lim_{\alpha\to 0} W_\alpha(P_{XY}\|Q_XQ_Y) &= \frac{\chi^2(P_{XY}\|Q_XQ_Y)}{2} = V_{\alpha,0,1-\frac{1}{\alpha},\frac{1}{\alpha^2},0}(P_{XY}\|Q_XQ_Y)\\
\lim_{\alpha\to 0} W_\alpha(P_X\|Q_X) &= \frac{\chi^2(P_X\|Q_X)}{2} = V_{\alpha,0,1-\frac{1}{\alpha},\frac{1}{\alpha^2},0}(P_X\|Q_X)\\
\lim_{\alpha\to 0} W_\alpha(P_Y\|Q_Y) &= \frac{\chi^2(P_Y\|Q_Y)}{2} = V_{\alpha,0,1-\frac{1}{\alpha},\frac{1}{\alpha^2},0}(P_Y\|Q_Y)
\end{aligned}\right\} \quad (**)$$

حال (*) و (**) با هم نتیجه می‌دهند که:

$$\boxed{\chi^2(P_{XY}\|Q_XQ_Y) \geq \chi^2(P_X\|Q_X) + \chi^2(P_Y\|Q_Y)}$$

Q.E.D. ■