۱. خواص Total Variation

۱. برای اثبات حکم این بخش، روی n استقرا می‌زنیم. ابتدا برای حالت $n=2$ حکم را اثبات می‌کنیم:

برای حالت $n=2$، فرض کنید دو متغیر تصادفی $X_1 \in \{x_1, \dots, x_n\}$ و $X_2 \in \{y_1, \dots, y_m\}$ داشته باشیم. X_1 با توزیع‌های p_1 و q_1

و X_2 با توزیع‌های p_2 و q_2 تولید می‌شوند. داریم:
$$d_{TV}(p_1 p_2, q_1 q_2) \leq d_{TV}(p_1, q_1) + d_{TV}(p_2, q_2)$$

$$\Longleftrightarrow \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} |p_{1,i} p_{2,j} - q_{1,i} q_{2,j}| \leq \frac{1}{2}\sum_{i=1}^{n}|p_{1,i} - q_{1,i}| + \frac{1}{2}\sum_{j=1}^{m}|p_{2,j} - q_{2,j}|$$

که به عنوان مثال $p_{1,i}$ احتمال $P_1(X_1 = x_i)$ است. یا $q_{2,j}$ همان $Q_2(X_2 = x_j)$ است.

برای اثبات حکم، از محدب بودن قدرمطلق استفاده می‌کنیم. با توجه به نامساوی مثلث داریم: $|ab - cd| \leq |a-c|b + |b-d|c$

با توجه به این نامساوی داریم: $|p_{1,i} p_{2,j} - q_{1,i} q_{2,j}| \leq |p_{1,i} - q_{1,i}| p_{2,j} + q_{1,i} |p_{2,j} - q_{2,j}|$

جمع می‌زنیم

$$\sum_{i=1}^{n}\sum_{j=1}^{m} |p_{1,i} p_{2,j} - q_{1,i} q_{2,j}| \leq \sum_{i=1}^{n}\sum_{j=1}^{m} |p_{1,i} - q_{1,i}| p_{2,j} + q_{1,i} |p_{2,j} - q_{2,j}| \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} |p_{1,i} - q_{1,i}| \underbrace{\sum_{j=1}^{m} p_{2,j}}_{1} + \underbrace{\sum_{i=1}^{n} q_{1,i}}_{1} \sum |p_{2,j} - q_{2,j}|$$

$$= \sum_{i=1}^{n} |p_{1,i} - q_{1,i}| + \sum_{j=1}^{m} |p_{2,j} - q_{2,j}|$$

پس حکم در حالت $n=2$ اثبات می‌شود.

Subject :　　　　　　　　　　　　　　　　　　　　　Date :

فرض و گام استقرا: فرض کنید حکم مسئله برای $n=k$ درست باشد، حال برای $n=k+1$ داریم:

$$\sum_{i=1}^{k+1} d_{TV}(p_i,q_i) = \underbrace{\sum_{i=1}^{k} d_{TV}(p_i,q_i)}_{\text{فرض استقرا}} + d_{TV}(p_{k+1},q_{k+1}) \geq d_{TV}\left(\prod_{i=1}^{k} p_i, \prod_{i=1}^{k} q_i\right) + d_{TV}(p_{k+1},q_{k+1})$$

حال برای نامساوی آخر، با توجه به $n=2$ به سادگی حکم زیر برقرار است:

$$d_{TV}\left(\prod_{i=1}^{k} p_i, \prod_{i=1}^{k} q_i\right) + d_{TV}(p_{k+1},q_{k+1}) \geq d_{TV}\left(\prod_{i=1}^{k+1} p_i, \prod_{i=1}^{k+1} q_i\right)$$

پس حکم مسئله برای $n=k+1$ نیز اثبات می‌شود و حکم مسئله اثبات شد.

۲. فرض می‌کنیم فضای حالات ما n حالت داشته باشد: $X \in \{x_1,\dots,x_n\}$. حال داریم که $g(x_i)=y_i$

دقت کنید که با توجه به یک به یک بودن g داریم که $i \neq j \Rightarrow y_i \neq y_j$. پس با توجه به این موضوع، g پوشا نیز است. (زیرا اگر نباشد: $\exists i \neq j : y_i = y_j$)

پس:
$$d_{TV}(P_X,q_X) = \frac{1}{2}\sum_{i=1}^{n} |P(x_i)-q(x_i)| = \frac{1}{2}\sum_{i=1}^{n} |P(g(x_i))-q(g(x_i))| = \frac{1}{2}\sum |P(y_i)-q(y_i)| = d_{TV}(P_Y,q_Y)$$

* دقت کنید که با توجه به یک به یکی و پوشایی g، y_iها در واقع جایگشت دیگری از x_iها هستند. پس برابری بالا صحیح است.

۳. فرض کنید $X \sim P_0, P_1$ و $Y \sim q$ داریم:

$$d_{TV}(P_0\otimes q, P_1\otimes q) = \frac{1}{2}\sum_{\substack{x\in X\\ y\in Y}} |P_0(x)q(y)-P_1(x)q(y)|$$

$$= \frac{1}{2}\sum_{\substack{x\in X\\ y\in Y}} |P_0(x)-P_1(x)|\, q(y) = \frac{1}{2}\sum_{x\in X} |P_0(x)-P_1(x)| \underbrace{\sum_{y\in Y} q(y)}_{1}$$

$$= \frac{1}{2}\sum_{x\in X} |P_0(x)-P_1(x)| = d_{TV}(P_0,P_1)$$

MICRO

٤. برای یک بعد داریم: $d_{TV}(\mathcal{N}(0,\sigma^2), \mathcal{N}(\mu,\sigma^2)) = \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\left|e^{-\frac{x^2}{2\sigma^2}} - e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right| dx$

$$= \frac{2}{2\sqrt{2\pi}\sigma}\int_{-\infty}^{\frac{\mu}{2}} \left(e^{-\frac{(x-\mu)^2}{2\sigma^2}} - e^{-\frac{x^2}{2\sigma^2}}\right) dx$$

با توجه به تقارن و $= 1 - \Phi\left(\frac{\mu}{2\sigma}\right) - \Phi\left(\frac{\mu}{2\sigma}\right) = 1 - 2\Phi\left(\frac{\mu}{2\sigma}\right)$

حال برای ابعاد بالاتر، داریم: $d_{TV}(\mathcal{N}(0,C), \mathcal{N}(\mu,C)) = d_{TV}(\mathcal{N}(0,I), \mathcal{N}(C^{-\frac{1}{2}}\mu, I))$

عملا ما با تبدیل $C^{-\frac{1}{2}}$، با توجه به یک به یکی این تبدیل (بخش دوم) به تبدیل خطی رسیدیم که فاصله این دو توزیع برابر مسئله است.

حال دقت کنید که برای یک بعد هم داشتیم: $d_{TV}(\mathcal{N}(0,\sigma^2), \mathcal{N}(\mu,\sigma^2)) = d_{TV}(\mathcal{N}(0,1), \mathcal{N}(\frac{\mu}{\sigma},1))$

که این تساوی هم با توجه به یک به یکی تبدیل بدیهی است. حال دقت کنید که $d_{TV}(\mathcal{N}(0,I), \mathcal{N}(C^{-\frac{1}{2}}\mu, I))$

عملا مولفه‌های این متغیرها که با هم Corr ندارند و Var همه آن‌ها ۱ است.

حال هدف ما تعریف q است که $q \otimes p$ برابر همان یک بعد میانگین $\|C^{-\frac{1}{2}}\mu\|$ شود. $\|\hat{\mu}\| = 1$ ، $C^{-\frac{1}{2}}\mu = \|C^{-\frac{1}{2}}\mu\|\hat{\mu}$

تعریف می‌کنیم: $q = \mathcal{N}(\|C^{-\frac{1}{2}}\mu\|\hat{\mu}, I)$ ، آنگاه با توجه به بخش دوم:

$$d_{TV}(\mathcal{N}(0,I), \mathcal{N}(\|C^{-\frac{1}{2}}\mu\|\hat{\mu}, I)) = d_{TV}(\mathcal{N}(0,1), \mathcal{N}(\|C^{-\frac{1}{2}}\mu\|_2, 1)) = 1 - 2\Phi\left(\frac{1}{2}\|C^{-\frac{1}{2}}\mu\|_2\right)$$

پس حکم ثابت شد.

۲. D_{KL} برای توزیع های گوسی

$$p \sim N(\mu_1, C_1) \quad , \quad q \sim N(\mu_2, C_2)$$

$$\Rightarrow D_{KL}(p \| q) = E_p\left[\log\left(\frac{p(x)}{q(x)}\right)\right] = E_p[\log p(x)] - E_p[\log q(x)]$$

برای $p(x)$ داریم:

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |C_1|}} \exp\left(-\frac{1}{2}(x-\mu_1)^T C_1^{-1}(x-\mu_1)\right)$$

$$\Rightarrow \log p(x) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(C_1)) - \frac{1}{2}(x-\mu_1)^T C_1^{-1}(x-\mu_1)$$

حال با توجه به ثابت بودن دو ترم اول در امید ریاضی، کافی است امید آخری را حساب کنیم:

$$E_p\left(-\frac{1}{2}(x-\mu_1)^T C_1^{-1}(x-\mu_1)\right) = -\frac{1}{2}E_p\left[(x-\mu_1)^T C_1^{-1}(x-\mu_1)\right] = -\frac{1}{2}E_p\left[tr\left((x-\mu_1)^T C_1^{-1}(x-\mu_1)\right)\right]$$

$$= -\frac{1}{2}E_p\left[tr\left(C_1^{-1}(x-\mu_1)(x-\mu_1)^T\right)\right] \xrightarrow{C_1 = C_1^T} -\frac{1}{2}E_p\left[C_1^{-1}\, tr\left((x-\mu_1)(x-\mu_1)^T\right)\right]$$

$$= -\frac{1}{2}E_p\left[C_1^{-1}\, \underbrace{tr\left((x-\mu_1)^T(x-\mu_1)\right)}_{E \to C_1}\right] = -\frac{1}{2}tr(C_1^{-1}C_1) = -\frac{k}{2}$$

برای q داریم به طور مشابه فقط ترم آخر آن را محاسبه می‌کنیم:

$$(x-\mu_2)^T C_2^{-1}(x-\mu_2) = (x-\mu_1+\mu_1-\mu_2)^T C_2^{-1}(x-\mu_1+\mu_1-\mu_2) = (x-\mu_1)^T C_2^{-1}(x-\mu_1) + 2(\mu_1-\mu_2)^T C_2^{-1}(x-\mu_1)$$
$$+ (\mu_1-\mu_2)^T C_2^{-1}(\mu_1-\mu_2)$$

حال برای E_p آن داریم:

$$E_p[(x-\mu_1)^T C_2^{-1}(x-\mu_1)] + \underbrace{2(\mu_1-\mu_2)^T C_2^{-1} E_p[x-\mu_1]}_{0} + (\mu_1-\mu_2)^T C_2^{-1}(\mu_1-\mu_2)$$

$$= tr(C_2^{-1}C_1) + tr(C_2^{-1}(\mu_1-\mu_2)^T(\mu_1-\mu_2)) = tr(C_2^{-1}(C_1+(\mu_1-\mu_2)^T(\mu_1-\mu_2)))$$

$$\Rightarrow D_{KL}(p\|q) = \frac{1}{2}\log\left(\frac{\det(C_2)}{\det(C_1)}\right) - \frac{k}{2} + \frac{1}{2}tr(C_2^{-1}(\mu_1-\mu_2)^T(\mu_1-\mu_2))$$

۳ آماره بسنده

۱. طرف اول: فرض می‌کنیم T آماره بسنده از X است:

$$P_{X|\theta}(X|\theta) = \underbrace{P_{X|T,\theta}(X|T=t(x),\theta)}_{\text{با تعریف بسنده بودن}}\, P_{T|\theta}(T=t(x)|\theta)$$

$$P_{X|T,\theta}(X|T,\theta) = P_{X|T}(X|T)$$

$$\Rightarrow P_{X|\theta}(X|\theta) = P_\theta(X) = \underbrace{P_{X|T}(X|T=t(x))}_{h(x)} * \underbrace{P_{X|\theta}(T,\theta)}_{g(t(x),\theta)}$$

پس طرف اول اثبات شد. طرف دوم: فرض کنیم داریم: $P_\theta(x) = g(t(x),\theta)\,h(x)$

$$\Rightarrow P_\theta(X|T=t(x)) = \frac{P_\theta(X=x, T=t(x))}{P_\theta(T=t(x))} =$$

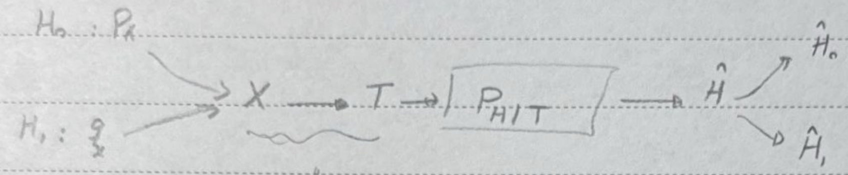
برای $P_\theta(T=t(x))$ داریم:

$$P_\theta(T=t(x)) = \sum_{\hat{x};T=t(\hat{x})} P_\theta(X, T(\hat{x})) = \sum_{\hat{x};T=t(\hat{x})} P_\theta(\hat{x}) = \sum_{\hat{x};T=t(\hat{x})} h(\hat{x})\,g(\theta,t(\hat{x}))$$

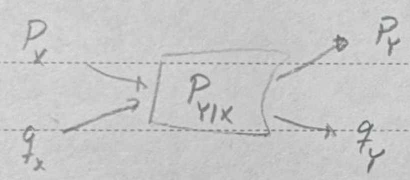
$$\Rightarrow P_\theta(X|T=t(x)) = \frac{P_\theta(X,T=t(x))}{\sum_{\hat{x}} P_\theta(\hat{x})} = \frac{h(x)\,g(t,\theta)}{\sum_{\hat{x}} h(\hat{x})\,g(t(\hat{x}),\theta)} = \frac{h(x)}{\sum_{\hat{x}} h(\hat{x})}$$

پس، $P_\theta(X \mid T=t(x)) = \dfrac{h(x)}{\sum_{\hat{x}} h(\hat{x})}$ چون وابسته به T نیست پس T بسنده است.

۲.



از طرفی می‌دانیم اگر $P_{X|Y}$ مستقل از X باشد پس Y آماره‌ای بسنده از X است:



$$D(P_{XY} \| q_{XY}) = E_{P_X}\left[\log \frac{P_X}{q_X}\right] + \underbrace{E_{P_X}\left[D(P_{Y|X} \| \overset{P_{Y|X}}{q_{Y|X}})\right]}_{0} = \underbrace{E_{P_Y}\left[\log\left(\frac{P_Y}{q_Y}\right)\right]}_{D(P_Y \| q_Y)} + \underbrace{E_{P_Y}\left[D(P_{X|Y} \| q_{X|Y})\right]}_{D(P_{X|Y} \| q_{X|Y} | P_Y)}$$

پس: $* \quad D(P_X \| q_X) = D(P_Y \| q_Y) + D(P_{X|Y} \| q_{X|Y} | P_Y)$

حال اگر T آماره‌ای بسنده باشد، $P_X^\theta P_{T|X} = P_T^\theta P_{X|T} \xleftarrow{\text{آحکم ۲ در مثال‌ها}} P_{X|Y} = P_{Y|X}$

پس ترم دوم تساوی در * صفر می‌شود و خواسته حکم برقرار است:

Y آماره‌ای بسنده از X $\rightarrow P_{X|Y} = q_{X|Y} \rightarrow D(P_X \| q_X) = D(P_Y \| q_Y) + \underbrace{D(P_{X|Y} \| q_{X|Y} | P_Y)}_{0} \rightarrow D(P_X \| q_X) = D(P_Y \| q_Y)$

پس یک طرف حکم ثابت می‌شود.

برای طرف دوم نیز داریم:

$$D(P_X \| q_X) = D(P_Y \| q_Y) \longrightarrow D(P_{X|Y} \| q_{X|Y}) = 0 \longrightarrow P_{X|Y} = q_{X|Y}$$

$$\downarrow$$

$X \perp \theta \mid Y \longrightarrow$ $\tilde{Y}$ آماره‌ی بسنده از X است

پس هر دو طرف اثبات شد: $D(P_X \| q_X) = D(P_T \| q_T) \Longleftrightarrow$ $\tilde{T}$ آماره‌ی بسنده از X است

۳. $$P_\theta(x) = \mathbb{1}(\theta=0) P_Y + \mathbb{1}(\theta=1) q_X = q_X \left(\frac{P_X}{q_X} \mathbb{1}(\theta=0) + \mathbb{1}(\theta=1) \right)$$

$$= \underbrace{q_X}_{\downarrow} \underbrace{\exp\Big(\log\big(\overbrace{\tfrac{P(x)}{q(x)}}^{T(x)} \mathbb{1}(\theta=0)\big) + \log\big(\mathbb{1}(\theta=1)\big)\Big)}_{\downarrow}$$

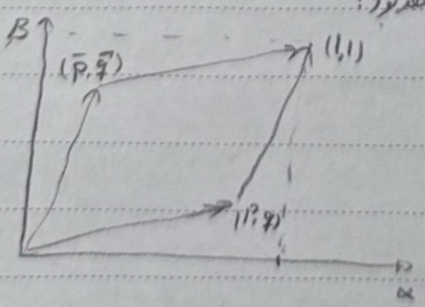
$$= h(x)\; g(T(x), \theta)$$

پس $P_\theta(x)$ به فرم بخش اول نوشته شد، پس $\tilde{T}$ آماره‌ی بسنده است.

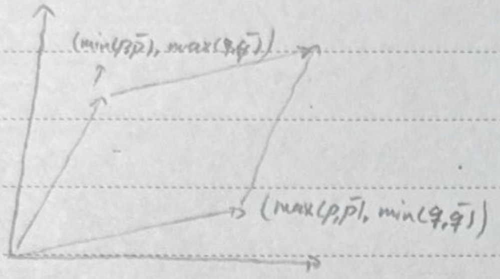
۴. $R(P,Q)$ برای توزیع های ساده

الف. $R(P,Q) = cl(co(R_{det}(P,Q))) \Rightarrow E = \{0\}, \{1\}, \{0,1\}$

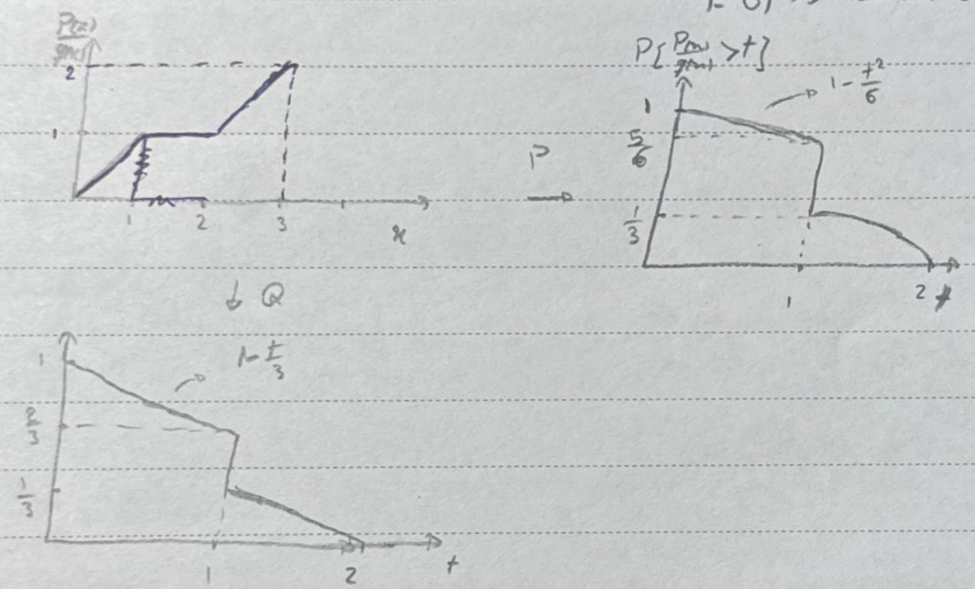
فرض کنید: $p > \frac{1}{2} > q$ با توجه به تقسیم‌ها، $4 = 2^2$ حالت خواهد بود:



حال به صورت کلی برای همه حالت‌های p و q به طور کلی داریم:



۲. ابتدا نمودارهای براساس Likelihood را رسم می‌کنیم:



حال کافی است با استفاده از لم Neyman-Pearson داریم که: $\alpha = P\{\log \frac{dP}{dQ} > \tau\} + \lambda P[\log \frac{dP}{dQ} = \tau]$

در بازه $0 \leq t \leq 1$ داریم: چون $\lambda = 0$ است: $\alpha = 1 - \frac{t^2}{6} \Leftarrow \beta = 1 - \frac{t}{3}$

پس با قرار دادن $t = 3(1-\beta)$ داریم: $\alpha = 1 - \frac{(1-\beta)^2 \times 3}{2}$

برای $t=1$ داریم: $\frac{1}{2} \leq \alpha \leq \frac{5}{6}$ از طرفی: $0 \leq \lambda \leq 1 \leftarrow \alpha = \frac{1}{2} + \frac{\lambda}{3}$, $\beta = \frac{1}{3} + \frac{\lambda}{3}$

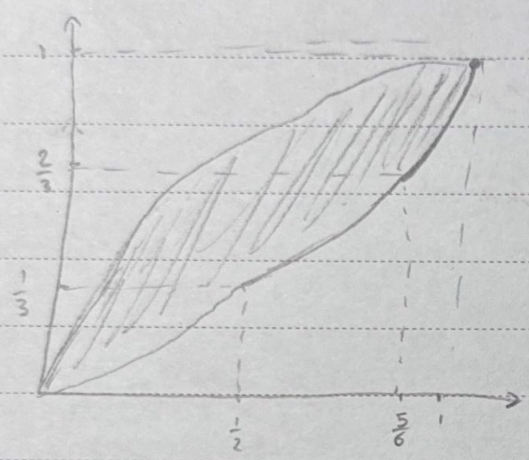
پس به طور کلی: $\alpha = \beta + \frac{1}{6}$ برای $t=1$

Subject : Date :

برای $1 < t \leq 2$ داریم: $\beta = \frac{1}{3} - \frac{t}{3}$, $\alpha = \frac{1}{2} - \left(\frac{t^2}{6} - \frac{t}{3}\right)$

$$\Rightarrow \alpha = \frac{2}{3} - \frac{3}{2}\beta^2$$

حال با رسم نمودار به صورت تقریبی داریم:



برای $\alpha = \frac{1}{2}$ داریم: $\beta_\alpha = \frac{1}{3}$

برای $\alpha = \frac{5}{6}$ داریم: $\beta_\alpha = \frac{2}{3}$

۳. داریم: $\log \frac{P(x)}{q(x)} = \frac{\mu^2}{2} - \mu x \geq \log t \Rightarrow x \leq \frac{\mu}{2} - \frac{\log t}{\mu}$

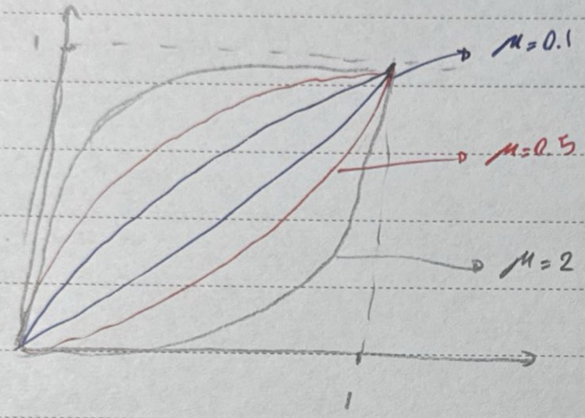
برای حل، $t>0$ ، $\max\{\alpha + t\beta : (\alpha, \beta) \in R(P,Q)\}$ را با بسی لینر:

$$\int_{\frac{P(x)}{Q(x)} > t}^{\infty} \frac{1}{\sqrt{2\pi}}\left(e^{-\frac{x^2}{2}} + t\, e^{-\frac{(x-\mu)^2}{2}}\right) dx = 1 - \Phi(t_\gamma) + t\,\Phi(\mu - t_\gamma) = \alpha^* + t\beta^*$$

داریم $\rightarrow t_\gamma = \frac{\mu}{2} - \frac{1}{\mu}\ln(t) \rightarrow \alpha^* + t\beta^* + 1 - \Phi(t_\gamma) + t\,\Phi(\mu - t_\gamma)$

$$\beta^* = \Phi(\mu - t_\gamma) \Rightarrow \alpha^* = 1 - \Phi(t_\gamma)$$

پس $R(P,Q)=\{(\alpha,\beta)=(1-\Phi(t),\Phi(\mu-t))\}$ $\Leftarrow$ $\beta_\alpha=\Phi(\mu-\Phi^{-1}(1-\alpha))$



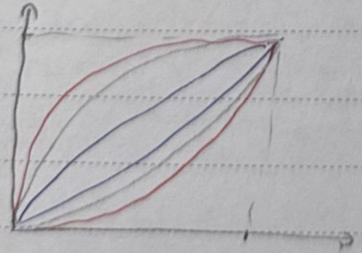
ج. $\log\frac{P^{n}(x)}{Q^{n}(x)}=\log\frac{\prod_{i=1}^{n}P(x_i)}{\prod_{i=1}^{n}Q(x_i)}=\sum\log\frac{P(x_i)}{Q(x_i)}=n\mu\left[\frac{\mu}{2}-\frac{\bar{X}}{n}\right]$

از تعریف لنیم: $T=\frac{\log P^{\otimes n}(x)}{\log Q^{\otimes n}(x)}$ ، به وضوح T آماره‌ای بسنده از x ها است. چون در سوال قبل آن را ثابت کردیم.

پس $n\mu\{\frac{\mu}{2}-\frac{\bar{X}}{n}\}$ نیز آماره‌ای بسنده خواهد بود و باز هم چون $\bar{X}$ را می‌توان از آن بدست آورد، $\bar{X}$ نیز بسنده خواهد بود.

پس با استفاده از حکم سوال قبل: $R(P^{\otimes n},Q^{\otimes n})=R(P_{\bar{X}},Q_{\bar{X}})$

حال داریم که: $P_{\bar{X}}\sim N(0,\frac{1}{n})$ و $Q_{\bar{X}}\sim N(\mu,\frac{1}{n})$ پس منحنی مانند سوال قبل می‌شود.



Subject : Date :

d_{TV} ، $R(P,Q)$ ـ ۵

$$\sup\{P(Z=0)-Q(Z=0)\}=\sup\Big\{\sum_x p(x)P(Z=0|x)-\sum_x q(x)P(Z=0|x)\Big\}$$ .1

$$=\sup\Big\{\sum_x P(Z=0|x)\,[p(x)-q(x)]\Big\}\longrightarrow P(Z=0|x)=\begin{cases}0 & p(x)\geq q(x)\\ 1 & p(x)<q(x)\end{cases}$$

$$\rightarrow \sup\{P(Z=0)-Q(Z=0)\}=\sum_{x:\,p(x)\,q(x)} p(x)-q(x)=d_{TV}(P,Q)$$

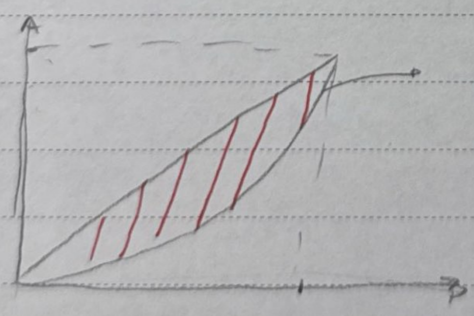
از طرفی ← $$\sup\{P(Z=0)-Q(Z=0)\}\xrightarrow{\alpha=P(Z=0)}\sup_{0\leq\alpha\leq 1}\ \sup_{\alpha=Q(Z=0)}\{\alpha-Q(Z=0)\}=$$

$$=\sup\{\alpha-\min_{\alpha=P(Z=0)}Q(Z=0)\}$$

$$=\sup\{\alpha-\beta_\alpha(P,Q)\}$$

$$\Rightarrow d_{TV}(P,Q)=\sup\{\alpha-\beta_\alpha(P,Q)\}$$

طبق این فرمول کافی است بزرگترین فاصله عمودی خط $\beta=\alpha$ را روی نمودار با β_α بگیریم، آن همان d_{TV} خواهد بود.



$d_{TV}=\max$ (خط های قرمز)

۲.

$$P_E = \inf_{P_{Z|X}} \left(\pi_0 \pi_{1|0} + \pi_1 \pi_{0|1}\right) = \inf_{P_{Z|X}} P(x) Z(1|x) \pi_0 + Q(x) Z(0|x) \pi_1$$

$$= \inf_{P_{Z|X}} \sum_x \left(P(x)\left(1 - Z(0|x)\right)\pi_0 + Z(0|x) q(x) \pi_1\right)$$

$$= \inf \left[\pi_0 - \sum Z(0|x)\left\{P(x)\pi_0 - q(x)\pi_1\right\}\right]$$

$$= \pi_0 - \sup\left[\sum_x Z(0|x)\left(p(x)\pi_0 - q(x)\pi_1\right)\right]$$

$$= \pi_0 - \sum_{x:\ \frac{P(x)}{q(x)} \geq \frac{\pi_1}{\pi_0}} \pi_0 P(x) - q(x)\pi_1$$

$$\rightsquigarrow Z(0|x) = \begin{cases} 1 & \pi_0 P(x) \geq \pi_1 q(x) \\ 0 & \pi_0 p(x) < \pi_1 q(x) \end{cases}$$

پس تست بهینه قطعی است، و با مقایسه نسبت $\frac{P(x)}{q(x)}$ با $\frac{\pi_1}{\pi_0}$ تعیین می‌شود.

۳. داشتیم که $T = \log \frac{P(x)}{q(x)}$ آماره‌ای بسنده است: پس $D_{KL}(P_X \| Q_X) = D_{KL}(P_T \| Q_T)$

از طرفی با توجه به hyp های supporting داریم: $P(Z=0) - \gamma Q(Z=0) \rightarrow \frac{1}{\gamma} = -\frac{d}{d\alpha}\beta_\alpha(P,Q) = \frac{dQ_X}{dP_X} = \frac{dP_T}{dQ_T}$

پس داریم:

$$D_{KL}(P_X \| Q_X) = D_{KL}(P_T \| Q_T) = E_T\left\{\log\left(\frac{dP_T}{dQ_T}\right)\right\} = \int_{-\infty}^{\infty} \log\left(\frac{dP_T}{dQ_T}\right)\frac{d\alpha}{dt}\, dt$$

$$= -\int_0^1 \log \frac{dP_X}{dQ_X}\, d\alpha = -\int_0^1 \log\left(-\frac{d}{d\alpha}\beta_\alpha(P,Q)\right) d\alpha$$

Subject : Date :

۶ ⊛ $\beta_{1-\epsilon}(P^{\otimes n}, Q^{\otimes n})$ با دقت بیشتر

طبق CLT داریم: $\dfrac{\frac{1}{n}F_n - D(P\|Q)}{\sqrt{V(P\|Q)}}\sqrt{n} \rightsquigarrow$

که F_n همان CDF متغیر تصادفی میانگین داده‌های ماست. از طرفی طبق قضیه Berry-esseen داریم:

$$\sup_z \left| P^n\left[\sqrt{n}\left(\frac{\frac{1}{n}F_n - D(P\|Q)}{\sqrt{V(P\|Q)}}\right) \leq z\right] - \Phi(z)\right| \leq \frac{C_0 k}{\sigma^3\sqrt{n}}$$

که در آن σ همان انحراف معیار یعنی $\sqrt{V(P\|Q)}$ و $k = E_P\left[\left|\log\left(\frac{P(x)}{Q(x)}\right) - D(P\|Q)\right|^3\right]$.

حال اگر قرار دهیم $Z = \sqrt{n}\left(\dfrac{\frac{1}{n}F_n - D(P\|Q)}{\sqrt{V(P\|Q)}}\right) \Leftarrow P(Z \geq \delta) = P(F_n \geq nD(P\|Q) + \sqrt{nV(P\|Q)}\,\delta)$

حال مقدار γ قرار میدهیم: $\log(\gamma) = nD(P\|Q) + \sqrt{nV(P\|Q)}\,\delta$ تا بتوانیم به گونه‌ای از قضیه 14.10 که ب (همان Strong convese) استفاده کنیم. طبق قضیه داریم:

$$\alpha_n - e^{nD(P\|Q) + \sqrt{nV(P\|Q)}\delta - nE} \leq P_{X^n}[F_n > \log\gamma]$$

$$\Rightarrow e^{n\left(D + \frac{\sqrt{V}}{\sqrt{n}}\delta - E\right)} \geq \alpha_n + P_{X^n}[Z \leq \delta] - 1$$

حال اگر δ را قرار دهیم: $\delta = \Phi^{-1}\left(\epsilon + \frac{Ck}{\sigma^3\sqrt{n}}\right)$ داریم:

Subject : Date :

$$e^{n\left(D+\sqrt{\frac{V}{n}}\delta-E\right)} \geqslant \alpha_n + \underbrace{P_{x^n}[Z\leqslant\delta]}_{\geqslant\varepsilon}-1 \geqslant \alpha_n+\varepsilon-1>0$$

$$\Rightarrow n\left(D+\sqrt{\frac{V}{n}}\delta-E\right)\geqslant 0$$

① $$\rightarrow \boxed{E\leqslant D(P\|Q)+\sqrt{\frac{V(P\|Q)}{n}}\,\phi^{-1}\left(\varepsilon+\frac{Ck}{\sigma^3\sqrt{n}}\right)}$$ ∎

از طرفی داریم $\rightarrow \alpha = P[F_n > n(D-\delta)] \Rightarrow \beta\leqslant e^{-n(D-\delta)}$

$$\Rightarrow P\left(Z>-\sqrt{\frac{n}{V}}\delta\right)=\alpha>1-\varepsilon$$

$$\rightarrow P\left(Z\leqslant\underbrace{-\sqrt{\frac{n}{V}}\delta}\right)<\varepsilon$$

$$\Rightarrow P\left(Z\leqslant-\sqrt{\frac{n}{V}}\,\phi^{-1}\left(\varepsilon+\frac{Ck}{\sigma^3\sqrt{n}}\right)\right)<\varepsilon$$

Berry-essen $\longrightarrow$ $$\beta\leqslant \exp\left(-n\left(D(P\|Q)+\sqrt{\frac{V}{n}}\,\phi^{-1}\left(\varepsilon+\frac{Ck}{\sigma^3\sqrt{n}}\right)\right)\right)$$

$$\Rightarrow \boxed{E\geqslant D+\sqrt{\frac{V}{n}}\,\phi^{-1}\left(\varepsilon-\frac{Ck}{\sigma^3\sqrt{n}}\right)}$$ ②

از ① و ② با توجه به مشتق‌پذیر بودن و بسط تیلور ϕ^{-1}: $\phi^{-1}\left(\varepsilon+\frac{c}{\sqrt{n}}\right)=\phi^{-1}(\varepsilon)+O(n^{-\frac{1}{2}})$ حکم نتیجه می‌شود.

۷. نگاه بیزی در خواص $R(P,Q)$

۱. داریم:

$$P_E = \pi_0 \pi_{1|0} + \pi_1 \pi_{0|1} \longrightarrow \min P_E = \min\left(\pi_0 \pi_{1|0} + \pi_1 \pi_{0|1}\right)$$

$$\rightarrow \min\left(\pi_0\left(1 - \pi_{0|0} + \frac{\pi_1}{\pi_0}\pi_{0|1}\right)\right) \xRightarrow{\text{equivalent}} \min \frac{\pi_1}{\pi_0} \underbrace{\pi_{0|1}}_{\beta} - \overbrace{\pi_{0|0}}^{\alpha} = \min \underbrace{\frac{\pi_1}{\pi_0}\beta - \alpha}_{\Delta}$$

پس:

$$\min_{P_{Z|X}} \left(\beta \frac{\pi_1}{\pi_0} - \alpha\right) = \min_{P_{Z|X}} \sum_x P(z|x)\left(\frac{\pi_1}{\pi_0} q(x) - p(x)\right) \rightarrow P(z|x) = \begin{cases} 1 & \frac{p(x)}{q(x)} \geq \frac{\pi_1}{\pi_0} \\ 0 & \frac{p(x)}{q(x)} < \frac{\pi_1}{\pi_0} \end{cases}$$

پس $\tau = \frac{\pi_1}{\pi_0}$.

۲. از قسمت قبل داریم: $\Delta = \beta \frac{\pi_1}{\pi_0} - \alpha$ $\leftarrow$ $\beta = \frac{\pi_0}{\pi_1}\alpha + \frac{\pi_0}{\pi_1}\Delta$ از طرفی می‌دانیم: $\beta_\alpha = \alpha^2$

$$\left.\begin{aligned} \beta_\alpha &= \alpha^2 \\ \beta &= \frac{\pi_0}{\pi_1}\alpha + \frac{\pi_0}{\pi_1}\Delta \end{aligned}\right\} \xrightarrow{\frac{\partial}{\partial \alpha}} \frac{\pi_0}{\pi_1} = 2\alpha^* \longrightarrow \beta^* = {\alpha^*}^2 \longrightarrow \beta^* = \frac{1}{4}\left(\frac{\pi_0}{\pi_1}\right)^2$$

$$\alpha^* = \frac{1}{2}\frac{\pi_0}{\pi_1}$$

۳. بجای لمای تساوی بیان شده باید در ازای فاصله بیان باشد یعنی خطای نوع II، دقت ما (α, β) باید کمتر از یک شود:

$$\alpha \leq 1 \longrightarrow \frac{\pi_0}{2\pi_1} \leq 1 \longrightarrow \pi_0 \leq 2\pi_1$$

بداهتاً برای β نیز این حد هم برقرار می‌شود.

Subject :   Date :

~~[illegible]~~ ۸ (*) به هم چسبیدگی!

۱. $P \triangleleft Q \iff A_n \in F_n$

داریم: $1-\alpha+\beta \geq \inf_{P_{Z|X}} \alpha+\beta = 1-d_{TV}(P_n, Q_n)$

$\rightarrow d_{TV}(P_n, Q_n) \geq \sup(P_n(A_n) - Q(A_n)) \geq P_n(A_n) - Q_n(A_n) \quad \forall A_n$

پس $\rightarrow \quad 1-\alpha+\beta \geq 1-P_n(A_n) + Q_n(A_n) \quad \forall A_n$

$\Rightarrow 1-\alpha+\beta \geq P_n(A_n^c) + Q_n(A_n)$

$P \triangleleft Q \Rightarrow P(A_n^c) \xrightarrow{n\to\infty} 1 \longrightarrow 1-\alpha+\beta \xrightarrow{n\to\infty} 1$ ✗

۲. $\|L_n\|^2 \langle L_n, L_n\rangle = \mathbb{E}_{X\sim Q_n}\left[\left(\frac{dP_n}{dQ_n}\right)^2\right] = \int \left(\frac{dP_n}{dQ_n}\right)^2 dQ_n(x) = \int \frac{dP_n}{dQ_n} dP_n = \mathbb{E}_{X\sim P_n}\left[\frac{dP_n}{dQ_n}\right]$

از فرض داریم: $P_n(A^n) = \int_{A^n} dP_n = \int_{A^n} L_n \, dQ_n = \mathbb{E}_{X\sim Q_n}[L_n] = \langle 1, L_n\rangle$

$\langle 1, L_n\rangle \leq \left(\int 1 \, dQ_n\right)^{\frac{1}{2}} \left(\underbrace{\int L_n^2 \, dQ_n}_{\|L_n\|^2}\right)^{\frac{1}{2}} \Rightarrow P_n(A^n) \leq \sqrt{Q_n(A^n)} \times c$

$c = \|L\|^2 < \infty$ (؟)

پس $P_n(A^n) \leq \sqrt{Q_n(A^n)} \cdot c \rightarrow P_n(A^n)^2 \leq Q_n(A^n) \cdot c^2$ چون اگر $Q_n(A^n) \to 0$ پس $P_n(A^n) \to 0$ (چون توان ۲ داره)!

پس $P \triangleleft Q$

MICRO

۲. ظرفیت کانال نقطه به نقطه

۱.

Message: $m \in \{1, \dots, N\}$, $N = 2^{nR}$

Enc $\rightarrow$ $X^n : \{1, \dots, N\}$

Dec $\rightarrow$ $Y^n : \{1, \dots, N\}$

پیام $\rightarrow$ $M \longrightarrow \boxed{X^n : \{1, \dots, N\} \rightarrow X^n} \xrightarrow{X^n(m)} \boxed{P_{Y|X}} \longrightarrow \boxed{D : Y^n \rightarrow \{1, \dots, N\}} \longrightarrow \hat{m}$

$$\rightarrow P[\hat{m} = m] = \frac{P_{Y|X}^{\otimes n}[y^n \mid X^n(m)]}{\sum_{m=1}^{N} P_{Y|X}^{\otimes n}(y^n \mid X^n(m))} \qquad (M \sim \text{Uniform}(N))$$

$$\rightarrow E_{X(1), \dots, X^n(N)}[\mathbb{P}[M = \hat{m}]] = E_{X^n(1) \dots X^n(N)}\Big[\sum_{y^n} \sum_{m} \underbrace{P_M(m)}_{\frac{1}{N}} P_{Y|X}^{\otimes n}(y^n \mid X^n(m)) P_{\hat{m}|Y^n}(m \mid y^n)\Big]$$

چون $P_M(m) = \frac{1}{N}$ بازای همه m ها، پس ثابت، به m مستقل است :

$$\Rightarrow E_{X(1) \dots X^n(N)}[\mathbb{P}[M = \hat{m}]] = \underbrace{N \times \frac{1}{N}}_{1} E_{X^n(1) \dots X^n(N)}\Big[\sum_{y^n} P_{Y|X}^{\otimes n}(y^n \mid X^n(1)) P_{\hat{m}|Y^n}(1 \mid y^n)\Big]$$

پس حکم اثبات شد

Subject :   Date :

۲. می دانیم: $P_{\hat{M}|Y^n}(1|y^n) = \dfrac{\frac{1}{N} P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)})}{\frac{1}{N}\sum_{m=1} P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(m)})}$

$$\Rightarrow E_{\hat{x}^n_{(1)},\dots,\hat{x}^n_{(N)}}\left[\mathbb{P}[m=\hat{m}]\right] = E_{\hat{x}^n_{(1)}\dots\hat{x}^n_{(N)}}\left[\sum_{y^n} P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)}) \frac{P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)})}{\sum_m P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(m)})}\right]$$

$$\underset{\text{Jensen}}{\geq} \sum_{y^n} E_{x_{(1)}}\left[\frac{P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)})^2}{P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)}) + E_{x_{(2)}\dots x_{(N)}}\left[\sum_{m=2}^{nR} P_{Y|X}(y^n|\hat{x}^n_{(m)})\right]}\right]$$

داریم: $E_{x_{(m)}}\left[P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(m)})\right] = P_Y^{\otimes n}(y^n)$

$$\text{پس} \Rightarrow E_{enc}\left[P[m=\hat{m}]\right] \geq \sum_{y^n} E_{x_{(1)}}\left[\frac{P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)})^2}{P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)}) + (2^{nR}-1) P_Y^{\otimes n}(y^n)}\right]$$

$$= \sum_{x^n, y^n} P_X^{\otimes n}(y^n, x^n) \frac{P_{Y|X}^{\otimes n}(y^n|x^n)}{P_{Y|X}^{\otimes n}(y^n|x^n) + (2^{nR}-1) P_Y^{\otimes n}(y^n)}$$

$$= E_{x, y^n}\left[\frac{P_{Y|X}^{\otimes n}[y^n|\hat{x}^n_{(1)}]}{P_{Y|X}^{\otimes n}(y^n|\hat{x}^n_{(1)}) + (2^{nR}-1) P_Y^{\otimes n}(y^n)}\right]$$

$$\geq E_{x, y^n}\left[\frac{1}{1 + 2^{nR} - 2^{-\log_2\left(\frac{P_{Y|X}^{\otimes n}(y^n|x^n)}{P_Y^{\otimes n}(y^n)}\right)}}\right] \checkmark$$

MICRO

Subject : Date :

۳. به آرگومان و تابع موجود در کسر قبل دقت کنید:

$$\log_2\left(\frac{P^{\otimes n}_{Y|X}(y^n|x^n)}{P^{\otimes n}_{Y^n}(y_i)}\right) = \sum_{i=1}^{n} \log_2\left(\frac{P_{Y|X}(y_i|x_i)}{P_Y(y_i)}\right) \simeq n\,E\left[\log_2\left(\frac{P_{Y|X}(y_i|x_i)}{P_Y(y_i)}\right)\right] = nI(X,Y)$$

$$\Rightarrow \quad E[P[m=\hat{m}]] \geq E_{X^n,Y^n}\left[\frac{1}{1+2^{n(R-I(X,Y))}}\right] \simeq \frac{1}{1+2^{n(R-I(X,Y))}}$$

حال اگر $I(X,Y) > R$ آن‌گاه،

$$\lim_{n\to\infty} \frac{1}{1+2^{n(R-I(X,Y))}} = \frac{1}{1+0} = 1$$

پس

$$\lim_{n\to\infty} \longrightarrow \quad E[P[m=\hat{m}]] \geq 1$$

از طرفی چون $E[P[m=\hat{m}]]$ امید ریاضی یک احتمال است، پس همه مؤلفه‌ها در نتیجه امید ریاضی آن همیشه کمتر از یک است:

$$1 \geq E[P[\hat{m}=m]]$$

پس طبق ساندویچ داریم: $\lim_{n\to\infty}: E[P[m=\hat{m}]]=1$

۱۰. ظرفیت کانال دسترسی چندگانه

۱. مانند سوال قبل، ما از کدگذاری تصادفی استفاده می‌کنیم، یعنی توزیع دلخواه P_X را روی نقطه‌ها انتخاب و با آن کار می‌کنیم.

$$m_X : \{1, \dots, 2^{nR_X}\} \longrightarrow X^n$$

$$m_Y : \{1, \dots, 2^{nR_Y}\} \longrightarrow Y^n$$

$$(X^n, Y^n) \longrightarrow \boxed{P_{Z|X,Y}} \longrightarrow Z^n \longrightarrow \boxed{Dec} \longrightarrow \hat{m} \; (m_X, m_Y)$$

این dec و enc را مانند سوال قبل می‌توان موازی انجام داد.

۲. مانند سوال قبل داریم:

$$P[\hat{m} = m] = \sum_{z,y} P[(\hat{m}_X, \hat{m}_Y) = (m_X, m_Y)] = \frac{P^{\otimes n}_{Z|X,Y}(Z \mid X^n(\hat{m}_X), Y^n(\hat{m}_Y))}{\sum_{z,y} P^{\otimes n}_{Z|X,Y}(Z \mid X^n(m_X), Y^n(m_Y))}$$

با توجه به:

$$P[\hat{m}_X = m_X] = \sum_{m_Y} P[(\hat{m}_X, \hat{m}_Y) = (m_X, m_Y)] \quad , \quad P[\hat{m}_Y = m_Y] = \sum_{m_X} P[(\hat{m}_X, \hat{m}_Y) = (m_X, m_Y)]$$

$$P_m(m) = P_{m_X}(m_X)\, P_{m_Y}(m_Y)$$

$$P[\hat{m} = m] = \sum_{m, Z^n} P_m(m)\, P_{Z|XY}(Z^n \mid X^n(m_X), X^n(m_Y))\, P_{\hat{m}|Z}(m \mid Z^n)$$

حال با توجه به تعریف:

$$E_{X(1) \dots X(2^{nR_X}), Y(1) \dots Y(2^{nR_Y})} \triangleq E_{XY}$$

مانند سوال قبل داریم:

Subject : Date :

$$\mathbb{E}_{XY}\left[P\{\hat{m}=m\}\right] = \mathbb{E}_{XY}\left[\sum_{z^n} P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)})\, P_{\hat{m}|z^n}(m|z^n)\right]$$

$$= \sum_{z^n} \mathbb{E}_{XY}\left[\frac{P_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)})\, P_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)})}{\sum_{m_x,m_y} P_{Z|X,Y}(z^n|X^n_{(m_x)},Y^n_{(m_y)})}\right]$$

$$= \sum_{z^n} \mathbb{E}_{XY}\left[\frac{P_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)})^2}{P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)}) + \sum_{m_x} P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(m_x)},Y^n_{(1)}) + \sum_{m_y} P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(m_y)}) + \sum_{\substack{m_x\neq 1\\ m_y\neq 1}} P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(m_x)},Y^n_{(m_y)})}\right]$$

Jensen
$$\geq \sum_{z^n} \mathbb{E}_{X(1),Y(1)}\left[\frac{P_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)})^2}{P_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)}) + \mathbb{E}_{X(2)\cdots Y(n)}\left[\sum_{m_x} P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(m_x)},Y^n_{(1)})\right] + \mathbb{E}_{X(2)\cdots Y(n)}\left[\sum_{m_y} P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(m_y)})\right] + \mathbb{E}_{X(2)\cdots Y(n)}\left[\sum_{\substack{m_x\neq 1\\ m_y\neq 1}}^{nR_x,nR_y} P_{Z|X,Y}(z^n|X^n_{(m_x)},Y^n_{(m_y)})\right]}\right]$$

$$\Rightarrow \mathbb{E}_{XY}(P\{m=\hat{m}\}) \geq \sum_{z^n} \mathbb{E}_{X(1)Y(1)}\left[\frac{P_{Z|X,Y}(z^n|X_{(1)},Y^n_{(1)})^2}{P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(1)},Y^n_{(1)}) + (2^{nR_x}-1)\, P^{\otimes n}_{Z|Y,Y}(z^n|Y^n_{(1)}) + (2^{nR_y}-1)\, P^{\otimes n}_{Z|X,Y}(z^n|X^n_{(1)}) + 2^{nR_x+nR_y} P_{Z|X,Y}(z^n)}\right]$$

با حذف اندیس‌های z' به توان‌ها و محاسبه امید ریاضی :

$$\cdots \geq E_{X^n,Y^n,Z^n}\left[\frac{P_{Z|XY}(z^n|x^n(1),y^n(1))}{P^{\otimes n}_{Z|X,Y}(z^n|x^n(1),y^n(1)) + 2^{nR_X} P^{\otimes n}_{Z|XY}(z^n|y^n(1)) + 2^{nR_Y} P^{\otimes n}_{Z|XY}(z^n|x^n(1)) + 2^{n(R_X+R_Y)} P^{\otimes n}_{Z|XY}(z^n)}\right]$$

$$\simeq E_{X^n,Y^n,Z^n}\left[\frac{P_{Z|XY}(z^n|x^n(1),y^n(1))}{1 + 2^{n(R_X - I(X;Z|Y))} + 2^{n(R_Y - I(Y;Z|X))} + 2^{n(R_X+R_Y - I(X,Y;Z))}}\right]$$

با توجه به فرض سؤال :

$$\lim_{n\to\infty} \longrightarrow E[P[\hat{m}=m]] \geq \frac{1}{1+0+0+0} = 1$$

پس در n های بزرگ $E[P[\hat{m}=m]] \geq 1$ ، پس با توجه به استدلال قبلی و بنا بر این می‌توان نوشت :

$$n\to\infty \longrightarrow E[P[m=\hat{m}]] = 1$$