نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

جلسه دهم

اثبات وارون در کانال دسترسی چندگانه

در جلسه قبل با استفاده از روش کدگشایی همزمان به ناحیه زیر رسیدیم:

$$\bigcup_{p(x_1),\, p(x_2)} \left\{ (R_1,R_2)\in\mathbb{R}^2: \quad \begin{array}{r} 0\le R_1\le I(X_1;Y|X_2),\\ 0\le R_2\le I(X_2;Y|X_1),\\ R_1+R_2\le I(X_1,X_2;Y). \end{array} \right\},$$

که در آن توزیع مشترک متغیرهای (X_1,X_2,Y) بشکل $p(x_1,x_2,y)=p(x_1)p(x_2)p(y|x_1,x_2)$ است که در آن $p(y|x_1,x_2)$ کانال دسترسی چندگانه مفروض است. یعنی باید روی تمامی ناحیه ها که با تغییر $p(x_1),p(x_2)$ به دست می آیند اجتماع بگیریم. برای محاسبه جملات اطلاعات متقابل در ناحیه فوق باید از توزیع مشترک متغیرهای (X_1,X_2,Y) بشکل $p(x_1,x_2,y)=p(x_1)p(x_2)p(y|x_1,x_2)$ استفاده کنیم که در آن $p(y|x_1,x_2)$ کانال دسترسی چندگانه مفروض است. برای محاسبه این اجتماع باید انتخابی از $p(x_1),p(x_2)$ را پیدا کنیم که تا جای ممکن مقادیر $I(X_1;Y|X_2)$، $I(X_2;Y|X_1)$ و $I(X_1,X_2;Y)$ را بیشینه کند. اما بیشینه کردن این جملات با همدیگر در تضاد هستند. مثلا فرض کنید که بخواهیم $I(X_1;Y|X_2)$ را بیشینه کنیم. ادعا میکنیم که این در صورتی اتفاق می افتد که X_2 مقدار ثابت داشته باشد زیرا اگر $p(x_1)$ و $p(x_2)$ توزیع دلخواهی باشد، آنوقت

$$\begin{aligned} I(X_1;Y|X_2) &= \sum_{x_2} I(X_1;Y|X_2=x_2)p_{X_2}(x_2)\\ &\le \max_{x_2} I(X_1;Y|X_2=x_2)\\ &\le \max_{x_2}\max_{p(x_1)} I(X_1;Y|X_2=x_2)\\ &= C_1. \end{aligned} \tag{۱}$$

که در آن رابطه (۱) به این دلیل درست است که میانگین وزن دار چند عدد (با وزن های نامنفی با مجموعه یک) از ماکزیمم آن اعداد بیشتر نیست. پس اگر X_2 با احتمال یک آن مقدار x_2 ای را انتخاب کند که $I(X_1;Y|X_2=x_2)$ را بیشینه میکند، آنوقت عبارت $I(X_1;Y|X_2)$ ماکزیمم میشود. پس در نتیجه $I(X_1;Y|X_2)$ در صورتی بیشینه میشود که X_2 مقدار ثابت داشته باشد. از طرف دیگر عبارت $I(X_2;Y|X_1)$ زمانی ماکزیمم میشود که X_1 مقدار ثابت داشته باشد. این دو با هم در تضاد هستند. پس ناحیه هایی که با تغییر $p(x_1),p(x_2)$ به دست می آیند، بشکل پنج ضلعی هایی خواهند بود که میتوانند کشیدگی در جهت های مختلف داشته باشند.

از آنجایی که میتوان با استفاده از اشتراک زمانی به خط واصل بین دو نقطه قابل حصول رسید، پوش محدب ناحیه کدگشایی همزمان نیز ناحیه ای قابل حصول است:

$$\text{CONV.HULL}\left(\bigcup_{p(x_1),\ p(x_2)}\left\{(R_1,R_2)\in\mathbb{R}^2:\begin{array}{r}0\le R_1\le I(X_1;Y|X_2),\\ 0\le R_2\le I(X_2;Y|X_1),\\ R_1+R_2\le I(X_1,X_2;Y).\end{array}\right.\right)\subseteq\mathcal{C}.$$

در این جلسه بخش وارون کانال دسترسی چندگانه را مورد بحث قرار میدهیم. خواهیم دید که ناحیه ظرفیت $\mathcal{C}$ همان ناحیه روش کدگشایی همزمان است. برای این کار باید نشان دهیم که هر زوج نرخ قابل حصول به ناحیه بالا تعلق دارد.

۱ پوش محدب اجتماع

پیش از ورود به بحث وارون میخواهیم مفهوم پوش محدب اجتماع را بررسی کنیم. برای سادگی فرض کنید که اجتماع تنها دو پنج ضلعی تعریف شده با دو زوج توزیع $p_1(x_1), p_1(x_2)$ و $p_2(x_1), p_2(x_2)$ را میگیریم. در این صورت دو ناحیه پنج ضلعی به شکل

$$\mathcal{R}_1=\left\{(R_1,R_2)\in\mathbb{R}^2:\begin{array}{r}0\le R_1\le I_{p_1}(X_1;Y|X_2),\\ 0\le R_2\le I_{p_1}(X_2;Y|X_1),\\ R_1+R_2\le I_{p_1}(X_1,X_2;Y).\end{array}\right\},$$

و

$$\mathcal{R}_2=\left\{(R_1,R_2)\in\mathbb{R}^2:\begin{array}{r}0\le R_1\le I_{p_2}(X_1;Y|X_2),\\ 0\le R_2\le I_{p_2}(X_2;Y|X_1),\\ R_1+R_2\le I_{p_2}(X_1,X_2;Y).\end{array}\right\},$$

خواهیم داشت. پوش محدب اجتماع $\mathcal{R}_1$ و $\mathcal{R}_2$ با در نظر گرفتن تمامی پاره خط هایی که نقاط $\mathcal{R}_1$ را به نقاط $\mathcal{R}_2$ وصل میکنند بدست می آید. این موضوع در شکل ۱ نشان داده شده است.

ابتدا فرض کنید که ما بجای کل پاره خط واصل نقاط دو مجموعه به نقاط وسط این پاره خط ها علاقه مند هستیم. این مجموعه را میتوان به شکل زیر نشان داد:

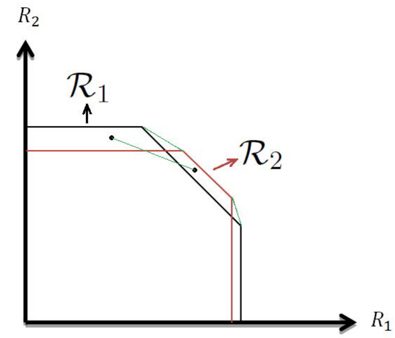
$$\begin{aligned}\mathcal{R}_3=\{(\tilde{R}_1,\tilde{R}_2)\in\mathbb{R}^2:\exists(R_1,R_2)\in\mathcal{R}_1,(R_1',R_2')\in\mathcal{R}_2:\\ (\tilde{R}_1,\tilde{R}_2)=\frac{1}{2}(R_1,R_2)+\frac{1}{2}(R_1',R_2')\}.\end{aligned}$$

تمرین ۱ نشان دهید که تعریف بالا معادل با تعریف زیر است که در آن $\oplus$ جمع مینکووسکی مجموعه ها است.

$$\mathcal{R}_3=\frac{1}{2}(\mathcal{R}_1\oplus\mathcal{R}_2).$$

قضیه ۲ $\mathcal{R}_3$ خود یک پنج ضلعی است که توسط معادلات زیر توصیف میشود.

$$\mathcal{R}_3=\left\{(R_1,R_2)\in\mathbb{R}^2:\begin{array}{r}0\le R_1\le \frac{1}{2}\big(I_{p_1}(X_1;Y|X_2)+I_{p_2}(X_1;Y|X_2)\big),\\ 0\le R_2\le \frac{1}{2}\big(I_{p_1}(X_2;Y|X_1)+I_{p_2}(X_2;Y|X_1)\big),\\ R_1+R_2\le \frac{1}{2}\big(I_{p_1}(X_1,X_2;Y)+I_{p_2}(X_1,X_2;Y)\big).\end{array}\right\}.$$



شکل ۱: اجتماع دو پنج ضلعی و پوش محدب آنها که با وصل کردن پاره خط های واصل بین نقاط این دو مجموعه به دست می آید.

اثبات: ابتدا نشان میدهیم که

$$\mathcal{R}_3 \subseteq \left\{ (R_1,R_2)\in\mathbb{R}^2: \begin{array}{r} 0\le R_1 \le \frac{1}{2}\big(I_{p_1}\big(X_1;Y|X_2\big)+I_{p_2}(X_1;Y|X_2)\big), \\ 0\le R_2 \le \frac{1}{2}\big(I_{p_1}\big(X_2;Y|X_1\big)+I_{p_2}(X_2;Y|X_1)\big), \\ R_1+R_2 \le \frac{1}{2}\big(I_{p_1}(X_1,X_2;Y)+I_{p_2}(X_1,X_2;Y)\big). \end{array} \right\}.$$

اگر $(\tilde{R}_1,\tilde{R}_2)\in\mathcal{R}_3$ آنوقت $(R_1,R_2)\in\mathcal{R}_1, (R_1',R_2')\in\mathcal{R}_2$ وجود دارند بطوریکه

$$\tilde{R}_1=\frac{1}{2}(R_1+R_1')$$

$$\tilde{R}_2=\frac{1}{2}(R_2+R_2')$$

از آنجایی که $(R_1,R_2)\in\mathcal{R}_1, (R_1',R_2')\in\mathcal{R}_2$ پس در نامساوی های مربوط به آن نواحی صدق میکنند. پس

$$\begin{aligned} \tilde{R}_1 &= \frac{1}{2}(R_1+R_1') \le \frac{1}{2}\big(I_{p_1}\big(X_1;Y|X_2\big)+I_{p_2}(X_1;Y|X_2)\big) \\ \tilde{R}_2 &= \frac{1}{2}(R_2+R_2') \le \frac{1}{2}\big(I_{p_1}\big(X_2;Y|X_1\big)+I_{p_2}(X_2;Y|X_1)\big) \\ \tilde{R}_1+\tilde{R}_2 &= \frac{1}{2}(R_1+R_2)+\frac{1}{2}(R_1'+R_2') \le \frac{1}{2}\big(I_{p_1}(X_1,X_2;Y)+I_{p_2}(X_1,X_2;Y)\big). \end{aligned}$$

پس نیز در نامساوی های مربوطه صدق میکند. پس اثبات این بخش کامل است. از طرف دیگر باید نشان دهیم

$$\left\{ (R_1,R_2)\in\mathbb{R}^2: \begin{array}{r} 0\le R_1 \le \frac{1}{2}\big(I_{p_1}\big(X_1;Y|X_2\big)+I_{p_2}(X_1;Y|X_2)\big), \\ 0\le R_2 \le \frac{1}{2}\big(I_{p_1}\big(X_2;Y|X_1\big)+I_{p_2}(X_2;Y|X_1)\big), \\ R_1+R_2 \le \frac{1}{2}\big(I_{p_1}(X_1,X_2;Y)+I_{p_2}(X_1,X_2;Y)\big). \end{array} \right\} \subseteq \mathcal{R}_3.$$

برای اثبات کافی است که ثابت کنیم که نقاط گوشه ای پنج ضلعی بالا در $\mathcal{R}_3$ هستند (چرا؟). پنج ضلعی بالا دو نقطه گوشه ای دارد که برابر هستند با (چرا؟)

$$\begin{bmatrix} \frac{1}{2}\big(I_{p_1}(X_1;Y)+I_{p_2}(X_1;Y)\big) \\ \frac{1}{2}\big(I_{p_1}\big(X_2;Y|X_1\big)+I_{p_2}(X_2;Y|X_1)\big) \end{bmatrix}$$

و

$$\begin{bmatrix} \frac{1}{2}\big(I_{p_1}(X_2;Y|X_1)+I_{p_2}(X_2;Y|X_1)\big) \\ \frac{1}{2}(I_{p_1}\big(X_1;Y)+I_{p_2}(X_1;Y)\big) \end{bmatrix}.$$

اما مثلا در مورد نقطه اول

$$\begin{bmatrix} \frac{1}{2}\big(I_{p_1}(X_1;Y)+I_{p_2}(X_1;Y)\big) \\ \frac{1}{2}(I_{p_1}\big(X_2;Y|X_1)+I_{p_2}(X_2;Y|X_1)\big) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} I_{p_1}(X_1;Y) \\ I_{p_1}\big(X_2;Y|X_1\big) \end{bmatrix} + \frac{1}{2}\begin{bmatrix} I_{p_2}(X_1;Y) \\ I_{p_2}\big(X_2;Y|X_1\big) \end{bmatrix}.$$

که نقطه وسط دو نقطه از $\mathcal{R}_1$ و $\mathcal{R}_2$ میباشد. اثبات کامل است. □

نکته ۳ در اثبات بالا در محاسبه نقاط گوشه ای از این نکته که $I(X_1X_2;Y) \leq I(X_1;Y|X_2) + I(X_2;Y|X_1)$ است استفاده کردیم. مساله ۴.۴ از کتاب الجمال و کیم ضرورت این شرط را نشان میدهد.

نکته ۴ در حالت کلی تر برای هر $\lambda \in [0,1]$ میتوان مجموعه زیر را تعریف کرد

$$\begin{aligned} \mathcal{R}_4 = \{(\tilde{R}_1,\tilde{R}_2) \in \mathbb{R}^2 : \exists (R_1,R_2) \in \mathcal{R}_1, (R_1',R_2') \in \mathcal{R}_2 : \\ (\tilde{R}_1,\tilde{R}_2) = \lambda(R_1,R_2) + (1-\lambda)(R_1',R_2')\}. \end{aligned}$$

استدلال مشابهی نشان میدهد که

$$\mathcal{R}_4 = \left\{ (R_1,R_2) \in \mathbb{R}^2 : \begin{array}{r} 0 \leq R_1 \leq \lambda I_{p_1}\big(X_1;Y|X_2\big) + (1-\lambda)I_{p_2}(X_1;Y|X_2), \\ 0 \leq R_2 \leq \lambda I_{p_1}\big(X_2;Y|X_1\big) + (1-\lambda)I_{p_2}(X_2;Y|X_1), \\ R_1 + R_2 \leq \lambda I_{p_1}(X_1,X_2;Y) + (1-\lambda)I_{p_2}(X_1,X_2;Y). \end{array} \right\}.$$

نکته ۵ در حالت کلی تر اگر بجای دو پنج ضلعی چندین پنج ضلعی $\mathcal{R}_i$ داشته باشیم که توسط زوج توزیع های $p_i(x_1), p_i(x_2), i = 1,2,\cdots k$ مشخص شده اند و همچنین وزن های نامنفی $\lambda_1, \lambda_2, \cdots, \lambda_k$ با جمع یک داشته باشیم آنوقت میتوان مجموعه زیر را تعریف کرد

$$\begin{aligned} \{(\tilde{R}_1,\tilde{R}_2) \in \mathbb{R}^2 : \exists (R_{1i},R_{2i}) \in \mathcal{R}_i : \\ (\tilde{R}_1,\tilde{R}_2) = \sum_i \lambda_i (R_{1i},R_{2i})\}. \end{aligned}$$

استدلال مشابهی نشان میدهد که این مجموعه برابر است با

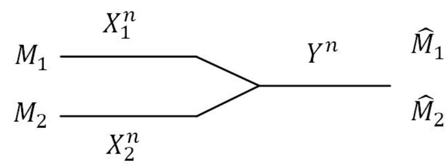
$$\left\{ (R_1,R_2) \in \mathbb{R}^2 : \begin{array}{r} 0 \leq R_1 \leq \sum_i \lambda_i I_{p_i}\big(X_1;Y|X_2\big), \\ 0 \leq R_2 \leq \sum_i \lambda_i I_{p_i}\big(X_2;Y|X_1\big), \\ R_1 + R_2 \leq \sum_i \lambda_i I_{p_i}(X_1,X_2;Y). \end{array} \right\}.$$

۲ اثبات وارون

یک زوج نرخ قابل حصول دلخواه $(R_1', R_2') \in \mathcal{C}$ را در نظر بگیرید. نشان می دهیم که

$$(R_1', R_2') \in \mathsf{CONV.HULL}\left(\bigcup_{p(x_1),\, p(x_2)} \left\{ (R_1, R_2) \in \mathbb{R}^2 : \begin{array}{c} 0 \le R_1 \le I(X_1; Y|X_2), \\ 0 \le R_2 \le I(X_2; Y|X_1), \\ R_1 + R_2 \le I(X_1, X_2; Y). \end{array} \right. \right).$$

از آنجایی که $(R_1', R_2') \in \mathcal{C}$ برای هر $\epsilon > 0$ یک کد برای این نرخ ها وجود دارد که ما را به احتمال خطای ϵ برساند. فرض کنید که این کد به طول n بوده و متغیرهای آن را بصورت شماتیک در شکل زیر نشان دهیم.



برای احتمال خطای ϵ باید $P(M_1 \ne \hat{M}_1) \le \epsilon$ و $P(M_2 \ne \hat{M}_2) \le \epsilon$ برقرار باشند. نشان میدهیم که

$$(R_1' - O(\epsilon), R_2' - O(\epsilon)) \in \mathsf{CONV.HULL}\left(\bigcup_{p(x_1)p(x_2)} \left\{ (R_1, R_2) : \begin{array}{c} 0 \le R_1 \le I(X_1; Y|X_2) \\ 0 \le R_2 \le I(X_2; Y|X_1) \\ R_1 + R_2 \le I(X_1, X_2; Y) \end{array} \right. \right).$$

که منظور از $O(\epsilon)$ در اینجا از مرتبه ϵ میباشد. با توجه به شکل ناحیه و با الهام گرفتن از قسمت وارون اثبات نقطه به نقطه میتوان انتظار داشت که روند اثبات باید به شکل زیر باشد.

$$n(R_1' - O(\epsilon)) = H(M_1) - O(n\epsilon) \le \cdots \le \sum_{i=1}^{n} I(X_{1i}; Y_i | X_{2i})$$

$$n(R_2' - O(\epsilon)) = H(M_2) - O(n\epsilon) \le \cdots \le \sum_{i=1}^{n} I(X_{2i}; Y_i | X_{1i})$$

$$n(R_1' + R_2' - O(\epsilon)) = H(M_1, M_2) - O(n\epsilon) \le \cdots \le \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i)$$

با توجه به روش اثبات نقطه به نقطه با استفاده از شکل n حرفی ناحیه، در اینجا نیز میتوان انتظار داشت که زنجیره نامساوی ها از شکل n حرفی عبارات ظاهر شده در ناحیه عبور کند.

$$n(R_1' - O(\epsilon)) = H(M_1) - O(n\epsilon) \le \cdots \le I(X_1^n; Y^n | X_2^n) \le \cdots \le \sum_{i=1}^{n} I(X_{1i}; Y_i | X_{2i})$$

$$n(R_2' - O(\epsilon)) = H(M_2) - O(n\epsilon) \le \cdots \le I(X_2^n; Y^n | X_1^n) \le \cdots \le \sum_{i=1}^{n} I(X_{2i}; Y_i | X_{1i})$$

$$n(R_1' + R_2' - O(\epsilon)) = H(M_1, M_2) - O(n\epsilon) \le \cdots \le I(X_1^n, X_2^n; Y^n) \le \cdots \le \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i)$$

باید جای نقطه چین ها را پر کنیم.

در اینجا تنها نامساوی اول را ثابت میکنیم و اثبات بقیه به عنوان تمرین به خواننده واگذار میشود. جهت اثبات $H(M_1) - O(n\epsilon) \leq \cdots \leq I(X_1^n; Y^n|X_2^n)$ بهتر است که از سمت راست شروع کنیم و به سمت چپ برسیم:

$$
\begin{aligned}
I(X_1^n; Y^n|X_2^n) &\geq I(M_1; \hat{M}_1|X_2^n) && (۲)\\
&= H(M_1|X_2^n) - H(M_1|\hat{M}_1, X_2^n)\\
&= H(M_1) - H(M_1|\hat{M}_1, X_2^n) && (۳)\\
&\geq H(M_1) - H(M_1|\hat{M}_1)\\
&= H(M_1) - nO(\epsilon)
\end{aligned}
$$

معادله (۲) بدلیل شکل شرطی نامساوی پردازش داده برقرار است زیرا $\hat{M}_1$ تابعی از Y^n است و برای هر مقدار $X_2^n = x_2^n$ داریم،

$$I(M_1; Y^n|X_1^n, X_2^n = x_2^n) = 0.$$

در نتیجه با مشروط کردن به هر مقدار $X_2^n = x_2^n$ زنجیره مارکف

$$M_1 - X_1^n - Y^n - \hat{M}_1 \ \ \big| X_2^n = x_2^n$$

را خواهیم داشت. معادله (۳) به علت استقلال پیام اول از ورودی دوم برقرار است. در مرحله آخر از نامساوی فانو استفاده کردیم. پس رابطه مورد نظر را ثابت کردیم.

جهت اثبات

$$I(X_1^n; Y^n|X_2^n) \leq \sum_{i=1}^{n} I(X_{1i}; Y_i|X_{2i})$$

سعی میکنیم که عینا اثبات نقطه به نقطه را تعمیم دهیم. یعنی آن اثبات را روبروی خود گذاشته و از روش بسط دادن جملات و نامساوی ها و یا تساوی هایی که آنجا داشتیم الگو برداری می کنیم.

$$
\begin{aligned}
I(X_1^n; Y^n|X_2^n) &= H(Y^n|X_2^n) - H(Y^n|X_1^nX_2^n)\\
&= H(Y_1|X_2^n) + H(Y_2|Y_1X_2^n) + \cdots + H(Y_n|Y_1\cdots Y_{n-1}X_2^n)\\
&\quad - H(Y_1|X_1^nX_2^n) - H(Y_2|Y_1X_1^nX_2^n) - \cdots - H(Y_n|Y_1\cdots Y_{n-1}X_1^nX_2^n)\\
&\leq H(Y_1|X_{21}) + H(Y_2|X_{22}) + \cdots + H(Y_n|X_{2n})\\
&\quad - H(Y_1|X_{11}X_{21}) - H(Y_2|X_{12}X_{22}) - \cdots - H(Y_n|X_{1n}X_{2n}) && (۴)\\
&= I(X_{11}; Y_1|X_{21}) + I(X_{12}; Y_2|X_{22}) + \cdots + I(X_{1n}; Y_n|X_{2n})
\end{aligned}
$$

معادله (۴) به این علت برقرار است که خروجی های کانال فقط به ورودی های کانال در همان زمان وابسته اند (نه به ورودی ها یا خروجی های قبلی کانال) پس

$$H(Y_i|Y_1\cdots Y_{i-1}X_1^nX_2^n) = H(Y_i|X_{1i}X_{2i}), \qquad i=1,2,\cdots,n.$$

همچنین برداشتن شرط ابهام را افزایش می دهد پس خواهیم داشت:

$$H(Y_i|Y_1\cdots Y_{i-1}X_2^n) \geq H(Y_i|X_{2i}), \qquad i=1,2,\cdots,n.$$

۱.۲ ترکیب n جمله در یک جمله

تا اینجا ثابت کردیم که

$$\begin{aligned}
R_1' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}),\\
R_2' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}),\\
R_1' + R_2' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i).
\end{aligned}$$

در کانال نقطه به نقطه اینگونه اثبات را ادامه میدادیم:

$$R - O(\epsilon) \leq \frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i) \leq \frac{1}{n}\cdot n \max_{p(x)} I(X;Y).$$

در اینجا هرکدام از جملات $I(X_i;Y_i)$ را با $\max_{p(x)} I(X;Y)$ کران بالا میزدیم. اگر بخواهیم مستقیما این کار را تقلید کنیم به نامساوی های زیر میرسیم:

$$\begin{aligned}
R_1' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) \leq \frac{1}{n}\cdot n \max_{p(x_1)p(x_2)} I(X_1;Y|X_2)\\
R_2' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}) \leq \frac{1}{n}\cdot n \max_{p(x_1)p(x_2)} I(X_2;Y|X_1)\\
R_1' + R_2' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) \leq \frac{1}{n}\cdot n \max_{p(x_1)p(x_2)} I(X_1,X_2;Y)
\end{aligned}$$

و یا

$$\begin{aligned}
R_1' - O(\epsilon) &\leq \max_{p(x_1)p(x_2)} I(X_1;Y|X_2)\\
R_2' - O(\epsilon) &\leq \max_{p(x_1)p(x_2)} I(X_2;Y|X_1)\\
R_1' + R_2' - O(\epsilon) &\leq \max_{p(x_1)p(x_2)} I(X_1,X_2;Y)
\end{aligned}$$

میرسیم. در اینجا جملات $I(X_{1i};Y_i|X_{2i})$ را با $\max_{p(x_1)p(x_2)} I(X_1;Y|X_2)$ و جملات $I(X_{2i};Y_i|X_{1i})$ را با $\max_{p(x_1)p(x_2)} I(X_2;Y|X_1)$ کران بالا زدیم. اما این کار خوب نیست. زیرا در ابتدای جلسه دیدیم که اگر $I(X_{1i};Y_i|X_{2i})$ در رابطه اول بخواهد به مقدار ماکزیممش برسد لازم است که X_{2i} یک مقدار ثابت باشد و اگر $I(X_{2i};Y_i|X_{1i})$ بخواهد به مقدار ماکزیممش برسد لازم است که X_{1i} مقدار ثابت باشد. این دو با هم در تضاد هستند و این تنش در کران بالای فوق رفع و رجوع نمی شود. در روابط فوق، به این نکته باید توجه داشت که توزیع مورد استفاده در $I(X_{1i};Y|X_{2i})$ همان توزیع استفاده شده در $I(X_{2i};Y|X_{1i})$ است. زمانی که یک باند بیرونی برای ناحیه ظرفیت می نویسیم، در تمام نامساوی های استفاده شده باید بررسی کنیم که آیا تساوی واقعاً می تواند بصورت همزمان در آن نامساوی ها اتفاق بیافتد یا خیر. در این حالت اگر به چند نامساوی رسیدیم که تساوی در آنها نمی تواند به طور همزمان اتفاق بیافتد، در واقع یک باند بیرونی ضعیف پیدا کرده ایم. در صورتی که از کران بالا استفاده کنیم به ناحیه بیرونی زیر میرسیم که ضعیف است (این اولین کران بیرونی ای بود که در ابتدای درس در موردش صحبت کردیم).

$$R_1' - O(\epsilon) \le \max_{p(x_1)p(x_2)} I(X_1;Y|X_2) = C_1$$
$$R_2' - O(\epsilon) \le \max_{p(x_1)p(x_2)} I(X_2;Y|X_1) = C_2$$
$$R_1' + R_2' - O(\epsilon) \le \max_{p(x_1)p(x_2)} I(X_1,X_2;Y) = C_{12}$$

پس نباید هر جمله را با مقدار ماکزیمم جایگزین کرد.

مجددا نامساوی های

$$R_1' - O(\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}),$$
$$R_2' - O(\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}),$$
$$R_1' + R_2' - O(\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i).$$

را در نظر بگیرید. یک تلاش دیگر این خواهد بود که سعی کنیم که اندیس i^* را بگونه ای پیدا کنیم که

$$\frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) \le I(X_{1i^*};Y_{i^*}|X_{2i^*})$$
$$\frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}) \le I(X_{2i^*};Y_{i^*}|X_{1i^*})$$
$$\frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) \le I(X_{1i^*},X_{2i^*};Y_{i^*})$$

اما این تلاش نیز بی نتیجه خواهد بود. به عنوان مثال اگر $n=2$ و

$$I(X_{11};Y|X_{21}) = 2, \qquad I(X_{21};Y|X_{11}) = 0,$$

$$I(X_{12};Y|X_{22})=0, \qquad I(X_{22};Y|X_{12})=2,$$

باشد، آنوقت خواهیم داشت

$$\frac{1}{2}\big(I(X_{11};Y|X_{21})+I(X_{12};Y|X_{22})\big)=1, \tag{۵}$$

$$\frac{1}{2}\big(I(X_{21};Y|X_{11})+I(X_{22};Y|X_{12})\big)=1. \tag{۶}$$

تحقیق کنید که در این حالت نمی توان یک مقدار مشخص $i^* \in \{1,2\}$ را پیدا کرد به گونه ای که

$$1 \le I(X_{1i^*};Y|X_{2i^*}),$$
$$1 \le I(X_{2i^*};Y|X_{1i^*}).$$

۲.۲ پوش محدب در اثبات وارون

تا اینجا ثابت کردیم که

$$R_1' - O(\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}), \tag{۷}$$

$$R_2' - O(\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}), \tag{۸}$$

$$R_1' + R_2' - O(\epsilon) \le \frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i). \tag{۹}$$

حال n توزیع مربوط به ورودی های کانال در لحظات مختلف $p(x_{1i}),p(x_{2i})$ را در نظر بگیرید و تعریف کنید

$$\mathcal{R}_i = \left\{ (R_1,R_2)\in\mathbb{R}^2 : \begin{array}{c} 0\le R_1 \le I(X_{1i};Y_i|X_{2i}), \\ 0\le R_2\le I(X_{2i};Y_i|X_{1i}), \\ R_1+R_2 \le I(X_{1i},X_{2i};Y_i). \end{array} \right\}, \qquad i=1,2,3,\cdots n.$$

همچنین وزن های $\lambda_i=\frac{1}{n}$ را در نظر بگیرید. طبق نکته ۵ هر نقطه $(R_1'-O(\epsilon),R_2'-O(\epsilon))$ که در نامساوی های (۷)-(۹) صدق کند قابل بیان بصورت ترکیب محدب نقاط $\mathcal{R}_i$ با وزن های $\lambda_i=\frac{1}{n}$ است. از آنجایی که

$$\mathcal{R}_i \subseteq \bigcup_{p(x_1),\,p(x_2)} \left\{ (R_1,R_2)\in\mathbb{R}^2 : \begin{array}{c} 0\le R_1\le I(X_1;Y|X_2), \\ 0\le R_2\le I(X_2;Y|X_1), \\ R_1+R_2\le I(X_1,X_2;Y). \end{array} \right\}.$$

میتوانیم نتیجه بگیریم که $(R_1'-O(\epsilon),R_2'-O(\epsilon))$ متعلق به

$$\text{CONV.HULL}\Bigg(\bigcup_{p(x_1),\,p(x_2)} \left\{ (R_1,R_2)\in\mathbb{R}^2 : \begin{array}{c} 0\le R_1\le I(X_1;Y|X_2), \\ 0\le R_2\le I(X_2;Y|X_1), \\ R_1+R_2\le I(X_1,X_2;Y). \end{array} \right. \Bigg)$$

میباشد. با میل دادن ϵ به سمت صفر اثبات کامل میشود.

۳ استفاده از متغیر کمکی برای بیان پوش محدب در ناحیه ظرفیت

میدانیم که یکی از معانی مشروط کردن به یک متغیر پوش محدب گیری است. در این بخش میخواهیم بیان معادلی از ناحیه ظرفیت را بیان کنیم که بجای پوش محدب گیری در آن از مشروط کردن به یک متغیر کمکی استفاده شده باشد. بصورت مشخص ادعا میکنیم کهناحیه ظرفیت را میتوان به شکل زیر بیان کرد

$$\bigcup_{p(x_1,x_2,q)=p(q)p(x_1|q)p(x_2|q)} \left\{ (R_1,R_2)\in\mathbb{R}^2: \begin{array}{r} 0\le R_1\le I(X_1;Y|X_2,Q), \\ 0\le R_2\le I(X_2;Y|X_1,Q), \\ R_1+R_2\le I(X_1,X_2;Y|Q). \end{array} \right\}$$

که در آن اجتماع روی تمامی متغیرهای Q, X_1, X_2 گرفته شده که در زنجیره مارکف $X_1 \to Q \to X_2$ صدق میکنند. این زنجیره مارکف توسط شرط $p(x_1,x_2,q)=p(q)p(x_1|q)p(x_2|q)$ در اجتماع گیری بالا نشان داده شده است. توزیع تمامی متغیرها بشکل $p(x_1,x_2,q,y)=p(x_1,x_2,q)p(y|x_1,x_2)$ خواهد بود که بیانگر زنجیره مارکف $Q\to X_1,X_2\to Y$ نیز هست.

یک توزیع مشخص $p(x_1,x_2,q)$ را در نظر بگیرید. دقت کنید که

$$I(X_1;Y|X_2,Q)=\sum_q p(Q=q)I(X_1;Y|X_2,Q=q)$$

و بصورت مشابه برای جملات دیگر. میخواهیم از نکته ۵ استفاده کنیم. فرض کنید که $\lambda_q=p(Q=q)$ را به عنوان یک وزن در نظر بگیریم. در این صورت جمع این وزن ها برابر یک است زیرا $\sum_q \lambda_q=\sum_q p(Q=q)=1$. همچنین توزیع X_1 و X_2 به شرط $Q=q$ را برای مقادیر مختلف q در نظر میگیریم. در این صورت پنج ضلعی متناظر با این توزیع برابر خواهد بود با

$$\left\{ (R_1,R_2)\in\mathbb{R}^2: \begin{array}{r} 0\le R_1\le I(X_1;Y|X_2,Q=q), \\ 0\le R_2\le I(X_2;Y|X_1,Q=q), \\ R_1+R_2\le I(X_1,X_2;Y|Q=q). \end{array} \right\}$$

اگر نقاط این پنج ضلعی ها برای مقادیر مختلف q را با وزن های λ_q ترکیب کنیم، مجموعه نقاطی بدست می آید که در معادلات زیر صدق میکنند:

$$\left\{ (R_1,R_2)\in\mathbb{R}^2: \begin{array}{r} 0\le R_1\le \sum_q \lambda_q I(X_1;Y|X_2,Q=q)=I(X_1;Y|X_2,Q), \\ 0\le R_2\le \sum_q \lambda_q I(X_2;Y|X_1,Q=q)=I(X_2;Y|X_1,Q), \\ R_1+R_2\le \sum_q \lambda_q I(X_1,X_2;Y|Q=q)=I(X_1,X_2;Y|Q). \end{array} \right\}$$

اگر وزن های λ_q دلخواه باشد، الفبای Q یعنی تعداد توزیع هایی که انتخاب میکنیم دلخواه باشد، و توزیع های شرطی X_1 و X_2 به شرط $Q=q$ هم دلخواه باشند، آنوقت ناحیه بدست آمده همان

$$\mathsf{CONV.HULL}\Bigg(\bigcup_{p(x_1),\, p(x_2)} \left\{ (R_1,R_2)\in\mathbb{R}^2: \begin{array}{r} 0\le R_1\le I(X_1;Y|X_2), \\ 0\le R_2\le I(X_2;Y|X_1), \\ R_1+R_2\le I(X_1,X_2;Y). \end{array} \right\}.$$

خواهد بود زیرا تمامی ترکیب محدب های مختلف بین نقاط اجتماع را لحاظ کرده ایم.

۱.۳ کامل کردن اثبات وارون با کمک ناحیه معادل

فرض کنید که بخواهیم مستقیما اثبات وارونی برای بیان ناحیه ظرفیت با متغیر کمکی بیابیم. هدف در اینجا بیان تکنیکی مرسوم در اثبات های وارون است. یک کد دلخواه در نظر بگیرید. دیدیم که میتوان با نوشتن زنجیره ای از نامساوی ها به نامساوی های

$$\begin{aligned} R_1' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}; Y_i|X_{2i}), \\ R_2' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2i}; Y_i|X_{1i}), \\ R_1' + R_2' - O(\epsilon) &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i). \end{aligned}$$

رسید. یک متغیر کمکی Q به این شکل میسازیم: فرض می کنیم متغیر Q مستقل از X^n و Y^n مربوط به کد باشد و به صورت یکنواخت اعداد مجموعه $\{1, 2, \cdots, n\}$ را به خود بگیرد. در این صورت اگر Q را به عنوان یک اندیس تصادفی در نظر بگیریم می توان نوشت:

$$\begin{aligned} I(X_{1Q}; Y_Q|X_{2Q}, Q) &= \sum_{q=1}^{n} p(Q = q) I(X_{1Q}; Y_Q|X_{2Q}, Q = q) \\ &= \sum_{q=1}^{n} \frac{1}{n} I(X_{1Q}; Y_Q|X_{2Q}, Q = q) && (۱۰) \\ &= \sum_{q=1}^{n} \frac{1}{n} I(X_{1q}; Y_q|X_{2q}, Q = q) \\ &= \sum_{q=1}^{n} \frac{1}{n} I(X_{1q}; Y_q|X_{2q}). && (۱۱) \end{aligned}$$

در رابطه فوق، تساوی (۱۰) بر اساس اینکه متغیر Q دارای توزیع یکنواخت است، یعنی $p(Q = q) = \frac{1}{n}$، نوشته شده است. توجه داشته باشید که در این معادله نمی توان Q را از رابطه بیرون انداخت و نوشت

$$I(X_{1Q}; Y_Q|X_{2Q}, Q = q) = I(X_{1Q}; Y_Q|X_{2Q}),$$

زیرا اگر چه Q مستقل از X^n و Y^n میباشد، اما X_{1Q} که ورودی اول در زمان Q است مستقل از Q نیست (زیرا دانستن Q ممکن است توزیع X_{1Q} را تغییر دهد). اما در معادله (۱۱) با توجه به فرض استقلال بین Q و X^n و Y^n ، می توان شرط $Q = q$ را از رابطه بیرون انداخت.

در نتیجه نامساوی های زیر را میتوانیم بنویسیم

$$\begin{aligned}
R_1' - O(\epsilon) &\le \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) = I(X_{1Q};Y_Q|X_{2Q},Q),\\
R_2' - O(\epsilon) &\le \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}) = I(X_{2Q};Y_Q|X_{1Q},Q),\\
R_1' + R_2' - O(\epsilon) &\le \frac{1}{n}\sum_{i=1}^{n} I(X_{1i},X_{2i};Y_i) = I(X_{1Q},X_{2Q};Y_Q|Q).
\end{aligned}$$

نکته جالب در روابط فوق این است که تابع توزیع احتمال بین Q، X_{1Q} ، X_{2Q} و Y_Q در هر سه نامساوی فوق، یک توزیع خاص و ثابت است. ادعا میکنیم که توزیع مشترک بین متغیرهای Q، X_{1Q} ، X_{2Q} و Y_Q را می توان به صورت زیر نوشت:

$$p(Q,X_{1Q},X_{2Q},Y_Q) = p(Q)p(X_{1Q},X_{2Q}|Q)p(Y_Q|X_{1Q},X_{2Q}).$$

اگر این موضوع را ثابت کنیم، میتوانیم نتیجه بگیریم که نقطه $(R_1' - O(\epsilon), R_2' - O(\epsilon))$ متعلق به

$$\bigcup_{p(x_1,x_2,q)=p(q)p(x_1|q)p(x_2|q)} \left\{ (R_1,R_2)\in\mathbb{R}^2 : \begin{array}{r} 0 \le R_1 \le I(X_1;Y|X_2,Q),\\ 0 \le R_2 \le I(X_2;Y|X_1,Q),\\ R_1+R_2 \le I(X_1,X_2;Y|Q). \end{array} \right\}$$

میباشد و اثبات تمام میشود. ابتدا توجه کنید که

$$I(X_{1Q},X_{2Q}|Q) = \sum_q p(Q=q)I(X_{1Q},X_{2Q}|Q=q) = \sum_q p(Q=q)I(X_{1q},X_{2q}) = 0.$$

در نتیجه

$$p(Q,X_{1Q},X_{2Q}) = p(Q)p(X_{1Q},X_{2Q}|Q).$$

سپس داریم

$$\begin{aligned}
p(Y_Q=y|X_{1Q}=x_1,X_{2Q}=x_2,Q=q) &= p(Y_q=y|X_{1q}=x_1,X_{2q}=x_2,Q=q)\\
&= p(Y_q=y|X_{1q}=x_1,X_{2q}=x_2)\\
&= q(y|x_1,x_2),
\end{aligned}$$

که ربطی به مقدار Q ندارد. پس

$$p(Y_Q=y|X_{1Q}=x_1,X_{2Q}=x_2,Q=q) = p(Y_Q=y|X_{1Q}=x_1,X_{2Q}=x_2), \qquad \forall x_1,x_2,y,q$$

در نتیجه زنجیره مارکف $Q - X_{1Q}X_{2Q} - Y_Q$ برقرار است. پس

$$p(Q,X_{1Q},X_{2Q},Y_Q) = p(Q,X_{1Q},X_{2Q})p(Y_Q|X_{1Q},X_{2Q})$$

و اثبات کامل است.