نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

جلسه پنجم: لم گنجایشی

لم گنجایشی محاسبات مربوط به محاسبه احتمال خطا در قسمت قابل حصول کانال نقطه به نقطه را تعمیم میدهد. این لم در آینده هنگام محاسبه احتمال خطاهای قسمت قابل حصول مسائل چندکاربره مورد استفاده قرار خواهد گرفت. در این جلسه ابتدا به مقدمات لم گنجایشی می پردازیم. سپس شکل کلی آن را بیان میکنیم.

۱ مقدمات

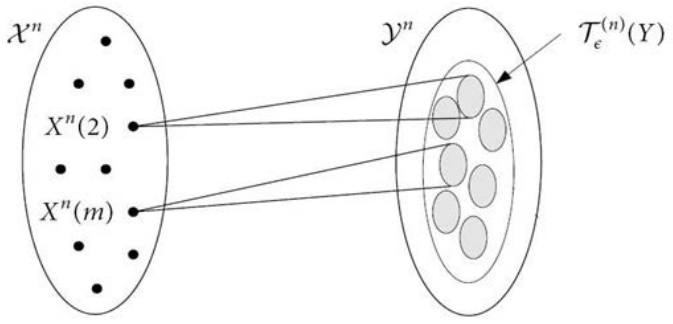
در مساله انتقال اطلاعات نقطه به نقطه یک توزیع ثابت $p(x)$ و یک کانال $p(y|x)$ را در نظر میگیریم. سپس 2^{nR} کلمه کد به طول n را به صورت $i.i.d$ از توزیع $p(x)$ میسازیم:

$$X^n(m) \sim \prod_{i=1}^{n} p(x_i), \qquad m \in [1 : 2^{nR}].$$

فرض کنید که کلمه کد اول $X^n(1)$ به عنوان پیام انتخاب شده و پس از عبور از کانال دنباله خروجی

$$\tilde{Y}^n \sim \prod_{i=1}^{n} p_Y(\tilde{y}_i)$$

را ایجاد کند. [۱] هدف نشان دادن این است که $\tilde{Y}^n$ با هیچکدام از $X^n(m), \quad m \in [2 : 2^{nR}]$ مشترکا نوعی نیست. یا به عبارت دیگر اگر دنباله های مشترکا نوعی با هرکدام از $X^n(m), \quad m \in [2 : 2^{nR}]$ را در نظر بگیریم، $\tilde{Y}^n$ جزو آنها نخواهد بود. هر کدام از کلمات کد $X^n(m), \quad m \in [2 : 2^{nR}]$ از کانال بی حافظه ی $p(y|x)$ گذشته و در مجموعه ی $\mathcal{Y}^n$ به یکی از $2^{nH(Y|X)}$ تا دنباله مشترکا نوعی y^n تناظر می یابد. این موضوع در شکل زیر نشان داده شده است.



[۱] دلیل اینکه خروجی را با $\tilde{Y}^n$ و نه Y^n نشان داده ایم بعدا آشکار میشود.

نقاط در طرف چپ شکل بالا $2^{nR}-1$ دنباله $X^n(m), \quad m \in [2:2^{nR}]$ از مجموعه ی $\mathcal{X}^n$ هستند. تعداد y^n های بطور شرطی نوعی با آن ها (مجموعه ی دوایر خاکستری) حداکثر

$$(2^{nR}-1)2^{nH(Y|X)} \leq 2^{n(R+H(Y|X))}$$

است. پس تا زمانی که $R < I(X;Y)$ باشد نسبت اندازه ی مجموعه ی دوایر خاکستری در شکل چپ به کل اندازه ی y^n های نوعی به صفر می رود.

$$\lim_{n\to\infty} \frac{2^{n(R+H(Y|X))}}{2^{nH(Y)}} = 0.$$

به عبارت دیگر اگر یک دنباله

$$\tilde{Y}^n \sim \prod_{i=1}^{n} p_Y(\tilde{y}_i)$$

بصورت تصادفی و یکنواخت از مجموعه نوعی $\mathcal{T}_\epsilon^{(n)}(p(y))$ انتخاب کنیم، احتمال اینکه در مجموعه ی دوایر خاکستری قرار بگیرد تقریبا صفر است.

یک نسخه اولیه از لم گنجایشی احتمال نوعی بودن $\tilde{Y}^n$ با یکی از $2^{nR}-1$ کلمه کد X^n بصورت تصادفی تولید شده را بررسی میکند. این لم بیان می دارد که در صورتی که $R < I(X;Y)$ این احتمال به سمت صفر می رود:

$$R < I(X;Y) \Rightarrow \lim_{n\to\infty} Prob[\exists\, 2 \leq m \leq 2^{nR} : (X^n(m), \tilde{Y}^n) \in \mathcal{T}_\epsilon^n(p(x,y))] \to 0. \qquad (۱)$$

یک اثبات شهودی از این موضوع به شرح زیر است: واقعه ی زیر را تعریف میکنیم:

$$E_m = \{(\tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(x,y))\}, \qquad 2 \leq m \leq 2^{nR}.$$

سپس از کران احتمال اجتماع استفاده میکنیم:[۲]

$$p(\bigcup_m E_m) \leq \sum_m p(E_m).$$

اما احتمال اینکه دنباله $\tilde{Y}^n$ با یک دنباله تصادفی X^n نوعی باشد تقریبا $2^{-n(I(X;Y))}$ است.

$$p((X^n, \tilde{Y}^n) \in \mathcal{T}_\epsilon^n) \leq 2^{-n(I(X;Y)-\delta)}.$$

پس

$$p(E_m) \leq 2^{-n(I(X;Y)-\delta)}.$$

سپس داریم:

$$\sum_m p(E_m) \leq 2^{nR} 2^{-n(I(X;Y)-\delta)} = 2^{-n(I(X;Y)-R-\delta)}.$$

همانطور که می بینیم اگر $R < I(X;Y)-\delta$ آنگاه احتمال $2^{-n(I(X;Y)-\delta)}$ وقتی $n \to \infty$ به سمت صفر می رود. اثبات کامل است.

[۲] اصولا هر وقت احتمال اجتماع یک سری وقایع را داشتیم و بدنبال کران بالا بودیم، نباید کران اجتماع فراموش شود

۲ لم گنجایشی

لم گنجایشی تعمیمی از رابطه (۱) میباشد. ما بصورت مرحله به مرحله این تعمیم را بیان خواهیم کرد تا به شکل نهایی این لم برسیم.

۱.۲ تعمیم اول: $\tilde{Y}^n$ از توزیع دلخواه

به عنوان اولین تعمیم فرض کنید که دنباله $\tilde{Y}^n$ لزوما خروجی کانال مربوط به ارسال $X^n(1)$ نیست، بلکه از توزیع دلخواهی آمده است و لزوما $i.i.d.$ نیست. دلیل اینکه از نمادگذاری $\tilde{Y}^n$ بجای Y^n در بخش قبل استفاده کردیم، مقدمه سازی برای همین تعمیم بود تا امکان تفاوت آن با خروجی کانال مشخص شود. پس فرض میکنیم که یک توزیع ورودی $p(x)$ و یک کانال $p(y|x)$ و یک دنباله $\tilde{Y}^n$ داشته باشیم بطوریکه توزیع $\tilde{Y}^n$ هیچ ارتباطی با توزیع $i.i.d.$ از $p(y) = \sum_x p(x)p(y|x)$ نداشته باشد. اما فرض میکنیم که مستقل از کلمات کد $X^n(m)$ تولید شده است. ثابت میکنیم که باز هم لم گنجایشی برقرار است:

$$R < I(X;Y) \Rightarrow \lim_{n\to\infty} Prob[\exists\, 1 \le m \le 2^{nR} : (X^n(m), \tilde{Y}^n) \in \mathcal{T}_\epsilon^n(p(x,y))] \to 0. \qquad (۲)$$

که در اینجا $m \in [1 : 2^{nR}]$ است. به اثبات بالا باز میگردیم. مجددا واقعه ی زیر را تعریف میکنیم:

$$E_m = \{(\tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(x,y))\}, \qquad 1 \le m \le 2^{nR}.$$

سپس از کران احتمال اجتماع استفاده میکنیم (اصولا هر وقت احتمال اجتماع یک سری وقایع را داشتیم و بدنبال کران بالا بودیم، نباید کران اجتماع فراموش شود):

$$p(\bigcup_m E_m) \le \sum_m p(E_m).$$

چون توزیع $\tilde{Y}^n$ دلخواه است و چیزی در مورد آن نمیدانیم، بهترین راه استفاده از احتمال شرطی است.[۳] داریم:

$$p(E_m) = \sum_{\tilde{y}^n} p(E_m|\tilde{Y}^n = \tilde{y}^n)p(\tilde{Y}^n = \tilde{y}^n)$$

توجه کنید که اگر $\tilde{y}^n$ نوعی نباشد، احتمال مشترکا نوعی بودن آن با هر دنباله x^n ای صفر است و E_m رخ نخواهد داد. پس تنها باید حالاتی را در نظر بگیریم که $\tilde{Y}^n$ یک دنباله نوعی بشود:

$$p(E_m) = \sum_{\tilde{y}^n \text{نوعی}} p(E_m|\tilde{Y}^n = \tilde{y}^n)p(\tilde{Y}^n = \tilde{y}^n).$$

اما احتمال اینکه دنباله نوعی ثابت و مشخص $\tilde{y}^n$ با یک دنباله تصادفی X^n نوعی باشد $2^{-n(I(X;Y)-\delta)}$ است. این موضوع را به طرق مختلف قبلا دیده ایم. آن را میتوان به شکل زیر هم مشاهده کرد: تعداد دنباله های x^n که با $\tilde{y}^n$ نوعی هستند

[۳] برخلاف تابع آنتروپی که مشروط کردن باعث تغییر آن میشود، مشروط کردن همیشه در احتمالات مجاز است و این یکی از نکات بسیار سودمند در مورد احتمال است.

برابر $2^{nH(X|Y)}$ است. اگر X^n تصادفی بصورت $i.i.d.$ انتخاب کنیم توزیع یکنواختی روی مجموعه نوعی خواهد داشت، و احتمال اینکه با $\tilde{y}^n$ نوعی شود برابر خواهد بود با $\frac{2^{nH(X|Y)}}{2^{nH(X)}} = 2^{-n(I(X;Y))}$. از آنجایی که این موضوع برای هر دنباله نوعی $\tilde{y}^n$ درست است،

$$\begin{aligned} p(E_m) &= \sum_{\tilde{y}^n \text{نوعی}} p(E_m|\tilde{Y}^n = \tilde{y}^n) p(\tilde{Y}^n = \tilde{y}^n) \\ &\leq \sum_{\tilde{y}^n \text{نوعی}} p(\tilde{Y}^n = \tilde{y}^n) 2^{-n(I(X;Y)-\delta)} \\ &\leq 2^{-n(I(X;Y)-\delta)}. \end{aligned}$$

ادامه اثبات کاملا مشابه قبل است.

۲.۲ تعمیم دوم: وابسته کردن انتخاب کلمات کد به همدیگر

در بخش قبل کلمات کد را بصورت مستقل از هم انتخاب میکردیم. اما میتوان روش های متنوعی از انتخاب کتاب کد بصورت تصادفی را متصور شد. مثلا فرض کنید که کلمات کد را اینگونه انتخاب کنیم. ابتدا $X^n(1)$ را بصورت $i.i.d.$ از $p(x)$ انتخاب میکنیم، سپس $X^n(2)$ را بصورت تصادفی از میان دنباله های نوعی که در فاصله ای خاص از $X^n(1)$ قرار دارند انتخاب میکنیم. سپس $X^n(3)$ را بصورت تصادفی از میان دنباله های نوعی که در فاصله ای خاص از $X^n(2)$ قرار دارند انتخاب میکنیم و الی آخر. نهایتا $\tilde{Y}^n$ را هم بصورت مستقل از $X^n(m), \quad m \in [1 : 2^{nR}]$ از یک توزیع دلخواه انتخاب میکنیم. سؤال این است که آیا رابطه (۲) همچنان برقرار است؟ باید به اثبات قبلی مراجعه کنیم و ببینیم که دقیقا به چه چیزی برای اثبات نیاز داشتیم. میبینیم که مرحله اصلی اثبات نوشتن رابطه

$$p(E_m) \leq 2^{-n(I(X;Y)-2\delta)}.$$

بود که برای درستی آن نیاز داریم که

- متغیرهای $\tilde{Y}^n$ و $X^n(m)$ برای هر $m \in [1 : 2^{nR}]$ دو به دو مستقل از هم باشند.
- توزیع حاشیه ای $X^n(m)$ بصورت $i.i.d.$ باشد.

پس مستقل بودن کلمات کد از همدیگر لازم نیست؛ تنها نیاز داریم که وقتی به توزیع حاشیه ای $X^n(m)$ نگاه میکنیم، توزیعی تقریبا یکنواخت روی مجموعه نوعی را مشاهده کنیم. مثلا این خاصیت در کتاب کدی که در ابتدای این زیربخش پیشنهاد دادیم، برقرار است. همچنین مشاهده شود که مستقل بودن مشترک تمامی کلمات کد و $\tilde{Y}^n$ لازم نیست؛ استقلال تک تک کلمات کد و $\tilde{Y}^n$ در نوشتن رابطه مورد نیاز کافی است.

نکته اصلی ای که ما را قادر به تعمیم های فوق کرد استفاده از کران اجتماع بود. کران اجتماع به ما این امکان را میدهد که تنها به استقلال دو به دو میان $\tilde{Y}^n$ و $X^n(m)$ برای مقادیر مختلف m نیاز داشته باشیم.

تمرین ۱ فرض کنید که دنباله ابتدا $X^n(1)$ را بصورت $i.i.d.$ از $p(x)$ انتخاب کنیم. سپس قرار دهیم:

$$X^n(2) = X^n(3) = \cdots = X^n(2^{nR}) = X^n(1).$$

نهایتا $\tilde{Y}^n$ را هم بصورت مستقل از $X^n(1)$ از یک توزیع دلخواه انتخاب میکنیم. آیا دو شرط لازم برای برقراری رابطه (۲) برآورده شده است؟ بطور مستقیم درستی این نتیجه را نیز تحقیق کنید.

تمرین ۲ فرض کنید که X یک متغیر دودویی باشد، $\mathcal{X} = \{0, 1\}$. و $p(x)$ یک توزیع یکنواخت روی $\mathcal{X}$ باشد. دنباله های $X^n(1), \cdots, X^n(2^{nR})$ را مستقل از هم و بصورت $i.i.d.$ از توزیع $p(x)$ انتخاب کرده ایم. سپس $\tilde{X}^n$ را با XOR کردن دو دنباله $X^n(1)$ و $X^n(2)$ میسازیم، و آن را از روی کانال $p(y|x)$ عبور داده تا $\tilde{Y}^n$ حاصل شود. آیا دو شرط لازم برای برقراری رابطه (۲) برآورده شده است؟

۳.۲ تعمیم سوم: کتاب کد ساختاریافته

بگذارید به مساله انتقال اطلاعات نقطه به نقطه باز گردیم: یک توزیع ثابت $p(x)$ و یک کانال $p(y|x)$ را در نظر میگیریم و سپس 2^{nR} کلمه کد به طول n را به صورت $i.i.d$ از توزیع $p(x)$ میسازیم:

$$X^n(m) \sim \prod_{i=1}^{n} p(x_i), \qquad m \in [1 : 2^{nR}].$$

در این روش ساختن کتاب کد، کلمات کد بصورت کاملا یکنواخت و تصادفی در مجموعه نوعی مربوط $\mathcal{T}(p(x))$ پخش خواهند شد و هیچ ساختار خاصی نخواهند داشت (بجز اینکه همه مربوط به یک مجموعه نوعی هستند). اما در مسائل با بیش از یک فرستنده و یک گیرنده نیاز به کدهای تصادفی ای داریم که دارای ساختار باشند و ارتباطی میان کلمات کد وجود داشته باشد. اما چگونه کدهای تصادفی دارای ساختار را تولید کنیم؟ یک راه مرتبط کردن تولید کلمات کد به همدیگر است که در بخش بالا مورد بحث قرار دادیم. این موضوع در تولید کتاب های تصادفی خطی کاربرد دارد. اما در مسائل شبکه به ساختارهای دیگری نیاز داریم که هم شکلی ساده داشته باشند و هم تحلیل آنها ساده باشد. در اینجا به یک روش متداول اشاره میکنیم.

فرض کنید که میخواهیم دو کلمه کد $x^n(1)$ و $x^n(2)$ را بگونه ای تولید کنیم که دارای مفهومی از "ارتباط" یا "نزدیکی" باشند. یک راه این است که یک دنباله مشخص u^n انتخاب کنیم و سپس از میان دنباله های x^n که با این دنباله مشترکا نوعی هستند، $x^n(1)$ و $x^n(2)$ را انتخاب کنیم. از آنجایی که این دو دنباله با یک دنباله مشترک u^n نوعی هستند، به هم نزدیک هستند. حال اگر بخواهیم که سری کلمه کد ساختار یافته بصورت تصادفی تولید کنیم، میتوانیم این کار را اینگونه انجام دهیم: یک توزیع دلخواه $p(u, x)$ را در نظر بگیرید. فرض کنید که یک دنباله $\tilde{U}^n$ از توزیع $i.i.d.$ از $p(u)$ را تولید کنیم. سپس با استفاده از همین یک دنباله $\tilde{U}^n$ به تعداد 2^{nR} بار از کانال $p(x|u)$ گذر داده تا کلمات کد حاصل شود: یعنی 2^{nR} کلمه کد را از توزیع $\prod_{i=1}^{n} p_{X|U}(x_i|\tilde{U}_i)$ بسازیم. به عبارت دیگر

$$X^n(m) \sim \prod_{i=1}^{n} p(x_i|\tilde{U}_i), \qquad m \in [1 : 2^{nR}].$$

در این صورت دنباله های $X^n(m)$ همگی با احتمال زیاد با دنباله $\tilde{U}^n$ نوعی خواهند شد. پس در این روش ساختن کتاب کد، کلمات کد بصورت کاملا یکنواخت و تصادفی در مجموعه دنباله های مشترکا نوعی با دنباله $\tilde{U}^n$ پخش خواهند شد. این دنباله ها با احتمال زیاد در مجموعه $\mathcal{T}(p(x))$ خواهند بود، اما هر دنباله $\mathcal{T}(p(x))$ با دنباله $\tilde{U}^n$ مشترکا نوعی نیست.

پس کلمات کد تولید شده ساختاریافته هستند از این جهت که وجه مشترکی میان آنها وجود دارد: همه آنها با دنباله $\tilde{U}^n$ مشترکا نوعی هستند.

اما لم گنجایشی برای کتاب های کد ساختاریافته به چه شکل خواهد بود؟ مشابه قبل فرض کنید که کلمه کد اول $X^n(1)$ به عنوان پیام انتخاب شده و پس از عبور از کانال دنباله خروجی $\tilde{Y}^n$ را ایجاد کند. هدف یافتن شرطی روی نرخ R است بطوریکه $\tilde{Y}^n$ با هیچکدام از $X^n(m), \quad m \in [2 : 2^{nR}]$ مشترکا نوعی نباشد. دقت کنید که در اینجا $\tilde{Y}^n$ و دنباله های $X^n(m), m \in [2 : 2^{nr}]$ همگی مشترکا نوعی با $\tilde{U}^n$ هستند. در واقع توزیع مشترک آنها بصورت زیر است:

$$p(\tilde{Y}^n = \tilde{y}^n, X^n(m) = x^n, \tilde{U}^n = \tilde{u}^n) = \prod_{i=1}^{n} p_U(\tilde{u}_i) \prod_{i=1}^{n} p_{Y|U}(\tilde{y}_i|\tilde{u}_i) \prod_{i=1}^{n} p_{X|U}(x_i|\tilde{u}_i).$$

یعنی $\tilde{Y}^n$ و $X^n(m)$ به شرط $\tilde{U}^n$ مستقل خواهند بود (زمانی که $\tilde{U}^n$ وجود نداشت مستقل بودند، اما در اینجا مستقل شرطی خواهند بود).

نشان خواهیم داد برای اینکه لم گنجایشی برقرار باشد بجای شرط $R < I(X;Y)$ نیاز به شرط $R < I(X;Y|U)$ داریم. نکته اصلی در آنالیز حالت قبل نتیجه زیر بود: اگر یک توزیع مشترک $p(x,y)$ داشته باشیم و دنباله های X^n و Y^n را بصورت $i.i.d.$ از توزیع حاشیه ای $p(x)$ و $p(y)$ تولید کنیم، آنوقت احتمال اینکه نوع مشترک آنها $p(x,y)$ بشود برابر است با $2^{-nI(X;Y)}$. دلیل این موضوع این است که ما دنباله ها را از توزیع $q(x,y) = p(x)p(y)$ تولید کرده ایم و میخواهیم نوعش $p(x,y)$ بشود و این احتمال برابر است با

$$2^{-nD\left(p(x,y)\|q(x,y)\right)} = 2^{-nD\left(p(x,y)\|p(x)p(y)\right)} = 2^{-nI(X;Y)}.$$

حال فرض کنید که دنباله های $X^n, Y^n, \tilde{U}^n$ را بصورت $i.i.d.$ از توزیع

$$q(u,x,y) = p_U(u)p_{X|U}(x|u)p_{Y|U}(y|u)$$

تولید کنیم، آنوقت احتمال اینکه نوع مشترک آنها[۴] $p(u,x,y)$ بشود برابر است با

$$2^{-nD\left(p(x,y,u)\|q(x,y,u)\right)} = 2^{-nD\left(p(x,y,u)\|p(u)p(x|u)p(y|u)\right)} = 2^{-nI_p(X;Y|U)}.$$

پس اگر E_m واقعه این باشد که سه تایی ام m $(X^n(m), \tilde{U}^n, \tilde{Y}^n)$ دارای نوع مشترک $p(u,x,y)$ باشد آنوقت با توجه به باند مجموع داریم:

$$P(\bigcup E_m) \leq \sum_{m=1}^{2^{nR}} p(E_m) = 2^{nR}2^{-nI_p(X;Y|U)} \to 0$$

اگر $R < I(X;Y|U)$.

تا اینجا فرض کرده بودیم که $\tilde{U}^n, \tilde{Y}^n$ بصورت $i.i.d.$ از $p(u,y)$ هستند. اما این فرض لازم نیست.

[۴] Joint Type

۴.۲ شکل نهایی لم گنجایشی

قضیه ۳ توزیع $p(u,x,y)$ روی متغیرهای (U,X,Y) را در نظر بگیرید. فرض کنید که دنباله های $\tilde{Y}^n$ و $\tilde{U}^n$ دارای توزیع دلخواه $p(\tilde{u}^n,\tilde{y}^n)$ باشند که لزوما بشکل $\prod_{i=1}^n p_{UY}(\tilde{u}_i,\tilde{y}_i)$ نیست. همچنین فرض کنید که 2^{nR} دنباله

$$X^n(m),\ \ 1\leq m\leq 2^{nR}$$

داشته باشیم بطوریکه توزیع حاشیه ای $X^n(m)$ و $\tilde{U}^n$ بصورت $\prod_{i=1}^n p_{X|U}(x_i|\tilde{u}_i)$ باشد:

$$p(X^n(m)=x^n|\tilde{U}^n=\tilde{u}^n)=\prod_{i=1}^n p_{X|U}(x_i|\tilde{u}_i).$$

بعلاوه زنجیره مارکف $X^n(m)-\tilde{U}^n-\tilde{Y}^n$ برقرار باشد. اما هیچ شرطی در مورد نحوه ارتباط $X^n(m)$ ها برای m های مختلف و $\tilde{U}^n,\tilde{Y}^n$ وجود ندارد. آنگاه اگر $R<I(X;Y|U)$ داریم:

$$\lim_{n\to\infty} Prob[\exists\ \ 1\leq m\leq 2^{nR}:(\tilde{U}^n,\tilde{Y}^n,X^n(m))\in\mathcal{T}_\epsilon^n(p(u,x,y))]=0.$$

نکته ۴ قضیه فوق بحث های قبلی را به دو طریق گسترش میدهد:

۱. کلمات کد $X^n(m),\quad m\in[1:2^{nR}]$ که بشرط $\tilde{U}^n$ مستقل از $\tilde{Y}^n$ هستند، لزوما مستقل از هم نیستند. یعنی حتی مثلا ممکن است که $X^n(2)=X^n(3)$ باشد زیرا تنها چیزی که لازم است شروطی مانند

$$X^n(2)-\tilde{U}^n-\tilde{Y}^n,\qquad X^n(3)-\tilde{U}^n-\tilde{Y}^n$$

میباشد که هیچ ربطی به نحوه ارتباط $X^n(2)$ و $X^n(3)$ با یکدیگر ندارند.

۲. دنباله های $\tilde{U}^n,\tilde{Y^n}$ توزیع دلخواه و نه لزوما $i.i.d.$ دارد. این شبیه بحثی است که در ابتدا در مورد کلی فرض کردن توزیع $\tilde{Y^n}$ داشتیم.

اثبات: ما در اینجا اساسا اثبات الجمال و کیم را دنبال خواهیم کرد. دو روش برای تکمیل انتهای اثبات وجود دارد که الجمال و کیم یک روش را برگزیده اند و ما جهت آموزنده بودن روش دیگری را انتخاب کرده ایم. اما ماهیت استدلالها یکسان است.

مشابه قبل به تعریف واقعه خطای زیر میپردازیم:

$$E_m=\{(\tilde{U}^n,\tilde{Y}^n,X^n(m))\in\mathcal{T}_\epsilon^n(p(u,y,x)\},\qquad 1\leq m\leq 2^{nR}$$

و سپس از کران احتمال اجتماع استفاده میکنیم:

$$p(\bigcup_m E_m)\leq\sum_m p(E_m).$$

حال داریم:

$$
\begin{aligned}
p(E_m) &= p((\tilde{U}^n, \tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x)) \\
&= \sum_{\tilde{u}^n, \tilde{y}^n} p\big((\tilde{U}^n, \tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x)) | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) p_{\tilde{U}^n, \tilde{Y}^n}(\tilde{u}^n, \tilde{y}^n) \\
&= \sum_{(\tilde{u}^n, \tilde{y}^n) \in \mathcal{T}_\epsilon^n(p(u,y))} p\big((\tilde{U}^n, \tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x)) | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) p_{\tilde{U}^n, \tilde{Y}^n}(\tilde{u}^n, \tilde{y}^n) \\
&\leq \max_{(\tilde{u}^n, \tilde{y}^n) \in \mathcal{T}_\epsilon^n(p(u,y))} p\big((\tilde{U}^n, \tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x)) | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big).
\end{aligned}
$$

در تساوی دوم از این نکته استفاده کردیم که اگر $(\tilde{u}^n, \tilde{y}^n)$ نوعی نباشد، احتمال مشترکا نوعی شدن آن با $X^n(m)$ صفر است. در مرحله آخر در واقع تصادفی بودن $\tilde{U}^n, \tilde{Y}^n$ را حذف کرده ایم و آن را بگونه ای در آورده ایم که انگار مقادیر ثابت $\tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n$ (که از لحاظ احتمال نوعی شدن بدترین حالت هستند) را با احتمال یک میگیرند. نشان میدهیم که برای هر $\tilde{u}^n, \tilde{y}^n$ نوعی دلخواه داریم:

$$
p\big((\tilde{U}^n, \tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x)) | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) \leq 2^{-n(I(X;Y|U)-\delta)}
$$

که ثابت میکند که $p(E_m) \leq 2^{-n(I(X;Y|U)-\delta)}$.

در اینجا باید احتمال اینکه $\big(\tilde{u}^n, \tilde{y}^n, X^n(m)\big)$ نوعی بشود را بیابیم که با توجه به زنجیره مارکف

$$
X^n(m) - \tilde{U}^n - \tilde{Y}^n
$$

توزیع $X^n(m)$ بصورت زیر است:

$$
p(X^n(m) = x^n | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n) = p(X^n(m) = x^n | \tilde{U}^n = \tilde{u}^n) = \prod_{i=1}^n p_{X|U}(x_i | \tilde{u}_i).
$$

در نتیجه

$$
\begin{aligned}
&p\big((\tilde{U}^n, \tilde{Y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x)) | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) \\
&\quad = p\big((\tilde{u}^n, \tilde{y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x)) | \tilde{U}^n = \tilde{u}^n\big) \\
&\quad = \sum_{x^n : (\tilde{u}^n, \tilde{y}^n, x^n) \in \mathcal{T}_\epsilon^n} p(X^n(m) = x^n | \tilde{U}^n = \tilde{u}^n)
\end{aligned}
$$

توجه کنید که برای هر دنباله $(\tilde{u}^n, \tilde{y}^n, x^n) \in \mathcal{T}_\epsilon^n$ داریم $(\tilde{u}^n, x^n) \in \mathcal{T}_\epsilon^n$ و در نتیجه

$$
p(X^n(m) = x^n | \tilde{U}^n = \tilde{u}^n) \approx 2^{-nH_p(X|U)}.
$$

اما چون به کران بالا نیاز داریم از رابطه زیر استفاده میکنیم:

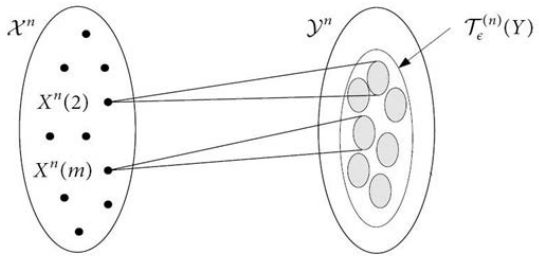
$$
p(X^n(m) = x^n | \tilde{U}^n = \tilde{u}^n) \leq 2^{-n(H_p(X|U)-\delta)}.
$$

پس

$$
\begin{aligned}
p\big((\tilde{u}^n, \tilde{y}^n, X^n(m)) \in \mathcal{T}_\epsilon^n(p(u,y,x))|\tilde{U}^n = \tilde{u}^n\big) &= \sum_{x^n:(\tilde{u}^n,\tilde{y}^n,x^n)\in\mathcal{T}_\epsilon^n} p(X^n(m) = x^n|\tilde{U}^n = \tilde{u}^n) \\
&\leq \sum_{x^n:(\tilde{u}^n,\tilde{y}^n,x^n)\in\mathcal{T}_\epsilon^n} 2^{-n(H_p(X|U)-\delta)} \\
&= 2^{-n(H_p(X|U)-\delta)} \sum_{x^n:(\tilde{u}^n,\tilde{y}^n,x^n)\in\mathcal{T}_\epsilon^n} 1 \\
&= 2^{-n(H_p(X|U)-\delta)} \big|\{\#x^n : (\tilde{u}^n, \tilde{y}^n, x^n) \in \mathcal{T}_\epsilon^n\}\big| \\
&\leq 2^{-n(H_p(X|U)-\delta)} 2^{n(H_p(X|U,Y)+\delta)} \\
&= 2^{-n(I_p(X;Y|U)-2\delta)}.
\end{aligned}
$$

□

نکته ۵ وجه تسمیه لم گنجایشی که ترجمه *Packing Lemma* میباشد به شرح زیر است: فرض کنید که 2^{nR} کلمه کد $X^n(m)$ را بصورت تصادفی انتخاب کنیم. هرکدام از این کلمات کد با $2^{nH(Y|X)}$ دنباله y^n مشترکا نوعی است. این مجموعه دنباله ها بصورت دوایر خاکستری رنگ در شکل زیر نشان داده شده اند.



با استفاده از لم گنجایشی نشان میدهیم که اگر $R < I(X;Y)$ باشد، میزان اشتراک این دوایر خاکستری رنگ تقریبا صفر خواهد بود (از آنجایی که در صورت دریافت دنباله هایی که با دو کلمه کد مشترکا نوعی هستند در کدگشایی دچار ابهام میشویم، تعداد چنین دنباله هایی باید محدود باشد، تا کل احتمال خطا به سمت صفر برود). پس نتیجه لم گنجایشی امکان گنجاندن 2^{nR} مجموعه های خاکستری رنگ در فضای $\mathcal{Y}^n$ است (در صورتی که $R < I(X;Y)$).

دلیل اینکه میزان اشتراک این دوایر خاکستری رنگ تقریبا صفر است، به شرح زیر است: یک کلمه کد خاص $X^n(1)$ را در نظر بگیرید. مجموعه دنباله های y^n مشترکا نوعی با $X^n(1)$ یک زیرمجموعه کاملا تصادفی از مجموعه نوعی $\mathcal{Y}^n$ است. دیدیم که نسبت اندازه ی مجموعه ی دوایر خاکستری مربوط به $X^n(2), \cdots, X^n(2^{nR})$ به اندازه مجموعه نوعی $\mathcal{Y}^n$ به صفر می رود. پس آن مجموعه ها درصد ناچیزی از زیرفضای نوعی را پر میکنند و احتمال اینکه با یک دایره خاکستری تصادفی جدید در فضای نوعی (مربوط به $X^n(1)$) برخورد داشته باشد، تقریبا صفر است. همچنین در صورتی که دوایر خاکستری تداخل زیادی با هم میداشتند، در کدگشایی با مشکل روبه رو میشدیم زیرا در صورت دریافت یک دنباله در خروجی که در اشتراک دو ناحیه خاکستری قرار میگرفت، در تشخیص اینکه کدام کلمه ورودی ارسال شده دچار ابهام میشدیم.