نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

جلسه هفتم: ادامه تابع نرخ - اعوجاج

در جلسه قبل مساله کدگذاری منبع با اتلاف یا مساله نرخ - اعوجاج را مطرح کردیم. در این جلسه فرمول تابع نرخ - اعوجاج $R(D)$ را می یابیم.

۱ فرمول تابع نرخ - اعوجاج

قضیه ۱ فرمول تابع نرخ - اعوجاج به شکل زیر توسط شانون بدست آمده است:

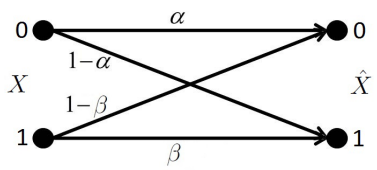
$$R(D) = \min_{p(\hat{x}|x):\ \mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}). \tag{۱}$$

به عبارت دیگر $R(D)$ مینیمم اطلاعات متقابل بین X و $\hat{X}$ روی تمام توزیع هایی از $p(x)p(\hat{x}|x)$ که شرط فاصله ی $\mathbb{E}[d(X,\hat{X})] \leq D$ را ارضا می کنند میباشد.

۱.۱ مفهوم رابطه تابع نرخ - اعوجاج

۱.۱.۱ برای منبع دودویی

رابطه $\mathbb{E}[d(X,\hat{X})] \leq D$ را میتوان به شکل $\sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D$ نیز نوشت. کانال $p(\hat{x}|x)$ که با عنوان "کانال تست"[۱] نیز شناخته میشود یک کانال با الفبای ورودی و خروجی یکسان $\mathcal{X}$ است. مثلا فرض کنید که X توزیع برنولی با احتمال $p(X=1)=p$ دارد. در این صورت کانال $p(\hat{x}|x)$ دارای ورودی دودویی و خروجی دودویی است که نمایش شماتیک آن در شکل زیر آمده است:



[۱] Test channel

مشاهده کنید که برای مشخص کردن این کانال باید دو پارامتر $\alpha, \beta \in [0,1]$ را مشخص کنیم. اما شرط متوسط اعوجاج برای این کانال با متر همینگ را میتوان بشکل زیر نوشت:

$$\begin{aligned}\mathbb{E}[d(X,\hat{X})] &= \sum_{x,\hat{x}} p_{X,\hat{X}}(x,\hat{x}) d(x,\hat{x}) \\ &= p_{X,\hat{X}}(0,1) + p_{X,\hat{X}}(1,0) \\ &= (1-p)(1-\alpha) + p(1-\beta).\end{aligned}$$

پس مینیمم گیری از اطلاعات متقابل X و $\hat{X}$ روی تمامی مقادیر α و β ای است که در شرط

$$(1-p)(1-\alpha) + p(1-\beta) \leq D$$

صدق میکنند. چند حالت خاص را در نظر بگیرید: اگر $D = 0$ باشد، آنوقت حتما می بایست $\alpha = \beta = 1$ باشد تا شرط

$$(1-p)(1-\alpha) + p(1-\beta) \leq 0$$

برقرار شود. در این حالت $\hat{X} = X$ بوده و $I(X;\hat{X})$ برابر با $H(X)$ خواهد بود. بازیابی با اعوجاج صفر همانند بازیابی بدون اتلاف نیازمند ارسال $H(X)$ بیت است.[۲] اما اگر فرض کنیم که $D \geq \min(p, 1-p)$ در جلسه قبل دیدیم که نیازی برای ارسال اطلاعات نیست $R = 0$. از فرمول بالا هم میتوان این موضوع را مشاهده کرد. فرض کنید که $p \leq 1-p$. در این صورت مینیمم گیری از اطلاعات متقابل X و $\hat{X}$ روی تمامی مقادیر α و β ای است که در شرط

$$(1-p)(1-\alpha) + p(1-\beta) \leq p$$

صدق میکنند. اگر انتخاب خاص $\beta = 0, \alpha = 1$ را در نظر بگیریم میبینیم که شرط برقرار شده است. اما در این صورت $\hat{X} = 0$ با احتمال یک برقرار بوده و اطلاعات متقابل X و $\hat{X}$ برابر صفر خواهد بود. این نشان میدهد که مقدار مینیمم

$$\min_{p(\hat{x}|x):\ \mathbb{E}[d(X,\hat{X})] \leq D} I(X;\hat{X})$$

برابر صفر است. نهایتا وقتی $0 < D < \min(p, 1-p)$ باشد، آنوقت مقدار

$$\min_{p(\hat{x}|x):\ \mathbb{E}[d(X,\hat{X})] \leq D} I(X;\hat{X})$$

عددی بین صفر و $H(X)$ خواهد بود.

[۲] البته توجه کنید که معیارهای بازیابی بدون اتلاف، و بازیابی با اتلاف با اعوجاج صفر دقیقا یکی نیستند. در یک کد برای بازیابی با اتلاف صفر با پارامتر (n, ϵ)، بطور متوسط ϵ درصد از بیت ها میتوانند تغییر کنند. اما در بازیابی بدون اتلاف با احتمال $1-\epsilon$ تمامی بیت ها سالمند و با احتمال ϵ حتی تمامی آنها میتوانند اشتباه باشند. شرط بازیابی با اتلاف ضعیف تر از شرط بازیابی بدون اتلاف است زیرا اگر با احتمال $1-\epsilon$ تمامی بیت ها سالم باشند، آنوقت متوسط اعوجاج برابر با ϵ ضربدر میزان اعوجاج حدکثری n، تقسیم بر n است که به ما متوسط اعوجاج ϵ بر بیت را میدهد.

۲.۱.۱ ارتباط با پخش شدگی کانال

کانال $p(\hat{x}|x)$ را میتوان به عنوان یک پخش شدگی از توزیع $p(x)$ در نظر گرفت. از آنجایی که الفبای $\hat{X}$ همان اندازه الفبای X است، پس تعداد نقاط پخش شده برابر اندازه الفبای X بوده و هرکدام از نقاط پخش شده متناظر با یک مقدار $\hat{X} = x$ خواهد بود. مثلا در حالت دودویی، دو نقطه پخش شدگی خواهیم داشت. اطلاعات متقابل $I(X;\hat{X})$ میزان متوسط پخش شدگی را نشان میدهد که سعی داریم آن را کمینه کنیم. پس شرط متوسط اعوجاج ما را از نزدیک کردن همزمان نقاط پخش شدگی به نقطه $p(x)$ منع میکند.

یکی از نقاط پخش شدگی، مثلا نقطه متناظر با $\hat{X} = 0$ را در نظر بگیرید. این نقطه پخش شدگی همان توزیع شرطی $p(x|\hat{X} = 0)$ میباشد، و بیانگر این است که اگر گیرنده $\hat{X} = 0$ را اعلام کرده باشد، چه میزان احتمال دارد که X ارسالی x باشد. متوسط اعوجاج در این حالت برابر با $\sum_x p(x|\hat{X} = 0)d(x,0)$ خواهد بود که با احتمال $p(\hat{X} = 0)$ اتفاق می افتد. عبارت $\sum_x p(x|\hat{X} = 0)d(x,0)$ یک ترکیب خطی با ضرایب نامنفی $d(x,0)$ از مولفه های نقطه پخش شدگی $p(x|\hat{X} = 0)$ است.

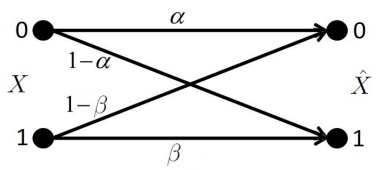
متوسط کل اعوجاج همان $\sum_{\hat{x}} p(\hat{X} = \hat{x}) \sum_x p(x|\hat{X} = \hat{x})d(x,\hat{x})$ میباشد که با عبارت بیان شده برای متوسط اعوجاج در بالا معادل است.

۳.۱.۱ محاسبه تابع نرخ - اعوجاج

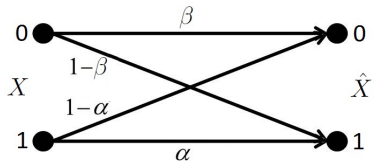
فرض کنید که بخواهیم مقدار دقیق تابع نرخ - اعوجاج را برای یک منبع دودویی با توزیع یکنواخت $p(X = 1) = \frac{1}{2}$ محاسبه کنیم. در این صورت مینیمم گیری از اطلاعات متقابل X و $\hat{X}$ روی تمامی مقادیر α و β ای است که در شرط

$$\frac{1}{2}(1-\alpha) + \frac{1}{2}(1-\beta) \leq D$$

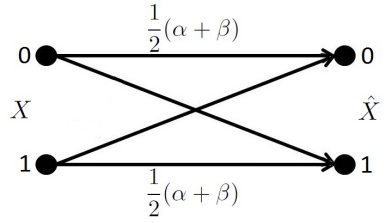
صدق میکنند. هدف پیدا کردن مقدار α و β است. بنابر تقارن انتظار داریم که حد اقل در یک مقدار $\alpha = \beta$ رخ دهد (که معادل با کانال تست باینری متقارن خواهد بود). اما چگونه میتوانیم این موضوع را ثابت کنیم؟ فرض کنید که یک مینیمم در $\alpha \neq \beta$ رخ دهد.



در این صورت اگر جای α و β را تغییر دهیم، هم یک مینیمم رخ میدهد.



حال کانال میانگین با $\alpha' = \beta' = \frac{1}{2}(\alpha + \beta)$ را در نظر بگیرید:



ادعا میکنیم که یک مینیمم نیز در $\alpha' = \beta'$ رخ میدهد. در ضمن شرط

$$\frac{1}{2}(1-\alpha') + \frac{1}{2}(1-\beta') \leq D$$

نیز برقرار است (چرا؟). اینکه یک مینیمم نیز در $\alpha' = \beta'$ رخ میدهد را میتوان از این نکته که اطلاعات متقابل تابعی محدب از کانال برای توزیع ورودی ثابت است، مشاهده کرد. این موضوع نتیجه میدهد که اطلاعات متقابل برای کانال میانگین از میانگین اطلاعات متقابل دو کانال (α, β) و (β, α) بیشتر نیست. در نتیجه یک مینیمم در کانال میانگین رخ میدهد.

برای یک اثبات مستقیم از این موضوع فرض کنید که یک سکه متعادل T را مستقل از X پرتاب کرده و بر اساس نتیجه آن تصمیم بگیریم که X را از کانال (α, β) و یا کانال (β, α) گذر دهیم. به عبارت دیگر فرض کنید که اگر $T = 1$ باشد، X را از کانال (α, β) گذر دهیم و خروجی را $\hat{X}_1$، و اگر $T = 2$ باشد، X را از کانال (β, α) گذر دهیم و خروجی را $\hat{X}_2$ بنامیم. در این صورت اگر نتیجه پرتاب سکه T را ندانیم، آنوقت کانال معادل از X به $\hat{X}_T$ همان کانال (α', β') خواهد بود. اما با دانستن T کانال ما به یکی از (α, β) و یا (β, α) تبدیل میشود زیرا

$$\begin{aligned}
&p(\hat{X}_T = \hat{x}|X = x) \\
&\quad = p(\hat{X}_T = \hat{x}|X = x, T = 0)p(T = 0|X = x) + p(\hat{X}_T = \hat{x}|X = x, T = 1)p(T = 1|X = x) \\
&\quad = p(\hat{X}_T = \hat{x}|X = x, T = 0)p(T = 0) + p(\hat{X}_T = \hat{x}|X = x, T = 1)p(T = 1) \\
&\quad = \frac{1}{2}p(\hat{X}_T = \hat{x}|X = x, T = 0) + \frac{1}{2}p(\hat{X}_T = \hat{x}|X = x, T = 1).
\end{aligned}$$

حال داریم:

$$\begin{aligned}
I(X; \hat{X}_T) &\leq I(X; \hat{X}_T, T) \\
&= I(X; T) + I(X; \hat{X}_T|T) \\
&= 0 + p(T = 1)I(X; \hat{X}_T|T = 1) + p(T = 2)I(X; \hat{X}_T|T = 2) \\
&= \frac{1}{2}I(X; \hat{X}_T|T = 1) + \frac{1}{2}I(X; \hat{X}_T|T = 2) \\
&= \frac{1}{2}I(X; \hat{X}_1) + \frac{1}{2}I(X; \hat{X}_2).
\end{aligned}$$

پس اثبات کامل است.

نکته ۲ در کتاب الجمال و کیم نشان داده شده است که حتی اگر X توزیع غیر یکنواخت داشته باشد، باز هم $\alpha = \beta$ اتفاق می افتد. اثبات را از آنجا دنبال کنید.

۲.۱ اثبات رابطه تابع نرخ - اعوجاج

این مساله دارای دو قسمت مستقیم و وارون است. در قسمت مستقیم باید نشان دهیم که با نرخ و اعوجاج داده شده میتوان منبع را بازسازی کرد. این موضوع نتیجه میدهد که نرخ بهینه کمتر مساوی نرخ داده شده است، یعنی

$$R(D) \leq \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}).$$

ملاحظه کنید که جهت نامساوی با آنچه در کدگذاری کانال داشتیم متفاوت است. دلیل این موضوع این است که اگر یک کد منبع خاص معرفی کنیم که با نرخ R به اعوجاج $D+\epsilon$ میرسد آنوقت میتوانیم از تعریف $R(D)$ در معادله

$$R(D) = \inf_{\text{قابل حصول } (R,D)} R,$$

نتیجه بگیریم که $R(D) \leq R$.

در قسمت وارون باید نشان دهیم که حداقل میزان نرخ برای بازسازی منبع لازم است، یعنی:

$$R(D) \geq \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}).$$

۱.۲.۱ قسمت مستقیم

قسمت مستقیم را در نظر بگیرید. ابتدا با اثبات یک نامساوی ساده تر شروع خواهیم کرد.

$$R(D) \leq \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} H(\hat{X})$$

از آن جایی که $H(\hat{X}) \geq I(X;\hat{X})$ ، پس

$$\min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}) \leq \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} H(\hat{X})$$

و نامساوی ای که میخواهیم ثابت کنیم درست، اما ضعیف است. یک توزیع دلخواه $p(\hat{x}|x)$ در نظر بگیرید بطوریکه $\mathbb{E}[d(X,\hat{X})] \leq D$. میخواهیم ببینیم به طور عملی چگونه میتوان با

$$R \geq H(\hat{X})$$

بیت، X^n را با اعوجاج کوچکتر مساوی D دریافت کرد. فرستنده X^n ای را که میخواهد ارسال کند را از کانال $\prod_{i=1}^{n} p(\hat{x}_i|x_i)$ عبور داده تا $\hat{X}^n$ ساخته شود. حال فرستنده میتواند $\hat{X}^n$ را با $nH(\hat{X})$ بیت بصورت بدون اتلاف فشرده کند و پس از فشرده سازی برای گیرنده ارسال کند (قادر به این کار هست زیرا $R \geq H(\hat{X})$). گیرنده اطلاعات فشرده شده را بازسازی کرده تا $\hat{X}^n$ را بدست آورد. در گیرنده $\hat{X}^n$ با احتمال خطای کم بازسازی میشود. برای سادگی محاسبات فرض کنید که بازسازی دقیق انجام شده است. در این صورت شرط

$$\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i)\big] \leq D+\epsilon$$

برقرار است چون فرستنده X^n را از کانال $\prod_{i=1}^n p(\hat{x}_i|x_i)$ عبور داده تا $\hat{X}^n$ ساخته شود در نتیجه $(\hat{X}^n, X^n)$ دارای توزیع $i.i.d.$ از $p(\hat{x}, x)$ هستند. پس

$$\begin{aligned}\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^n d(X_i, \hat{X}_i)\big] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[d(X_i, \hat{X}_i)\big] \\ &\leq \frac{1}{n}\sum_{i=1}^n D \\ &= D.\end{aligned}$$

پس شکل ضعیفتر را ثابت کردیم.

حال برای اینکه شکل قوی تر را ثابت کنیم از خود می پرسیم که کجای اثبات ما ضعیف بود و چگونه می توان آن را بهبود بخشید؟ به نظر می آید قسمتی که ضعیف است نویزی است که در فرستنده وارد مساله می شود. زمانی که فرستنده X^n ای را که می خواهد ارسال کند را از کانال $\prod_{i=1}^n p(\hat{x}_i|x_i)$ عبور داده تا $\hat{X}^n$ ساخته شود. در حین عبور نویزی وارد شده که سپس برای فشرده سازی آن $nH(\hat{X})$ بیت صرف کرده ایم. در حقیقت مقداری از nR بیت ارسالی صرف فشرده سازی نویز اضافه شده می شود. اما میزان نویز اضافه شده به مساله چقدر است؟ شهودا جهت شبیه سازی n استفاده از کانال $\prod_{i=1}^n p(\hat{x}_i|x_i)$ نیاز به $nH(\hat{X}|X)$ بیت تصادفی داریم. پس میتوان امید داشت که پس از حذف بخش نویز از بیت های فشرده شده به نرخ $nH(\hat{X}) - nH(\hat{X}|X) = nI(\hat{X}; X)$ رسید. اگر به عبارت (۱) نگاه کنیم می بینیم که بجای $H(\hat{X})$ دقیقا همین مقدار $I(X; \hat{X})$ را داریم که در واقع میزان اطلاعاتی است که $\hat{X}$ از X دارد که منطقی هم به نظر می رسد.

چگونه میتوان به $I(X; \hat{X})$ دست یافت؟ برای دستیابی به لم پوششی با $U = \emptyset$ نیاز داریم. اثبات دقیق به شکل زیر است:

یک توزیع دلخواه $p(\hat{x}|x)$ در نظر بگیرید بطوریکه $\mathbb{E}[d(X, \hat{X})] \leq \frac{D}{1+\epsilon}$. سپس کدی به این شکل میسازیم:

۱. ساختار کتاب کد: فرض کنید که 2^{nR} دنباله $\hat{X}^n(1), \hat{X}^n(2), \cdots, \hat{X}^n(2^{nR})$ را با استفاده از توزیع $i.i.d.$ $p(\hat{x})$ به عنوان کتاب کد تولید کنیم و بین فرستنده و گیرنده به اشتراک بگذاریم.

۲. کدگذار پس از مشاهده دنباله منبع X^n تولید شده از $p(x)$ بصورت $i.i.d.$، بدنبال $m \in [1 : 2^{nR}]$ ای میگردد بطوریکه X^n و $\hat{X}^n(m)$ مشترکا نوعی باشند. کدگذار اندیس m را بعنوان خروجی کدگذار اعلام میکند. اگر چند m با این خاصیت موجود باشد، کدگذار یکی را بدلخواه انتخاب میکند و اگر چنین m ای موجود نباشد، کدگذار $m = 1$ را برای گیرنده ارسال میکند.

۳. کدگشا پس از دریافت m، دنباله $\hat{X}^n(m)$ را به عنوان بازسازی منبع اعلام میکند.

آنالیز میزان اعوجاج بدست آمده: طبق لم پوششی اگر $R > I(X; \hat{X})$ باشد با احتمال بالا دنباله X^n با حداقل یکی از این دنباله ها نوعی خواهد بود و کدگذاری ممکن است. فرض کنید که E متغیر نشانگر این واقعه باشد که کدگذار نتواند یک دنباله مشترکا نوعی پیدا کند. فرض کنید که n را به اندازه کافی بزرگ انتخاب کرده ایم که $p(E = 1) \geq 1 - \epsilon$.

سپس داریم:

$$\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i)\big] = p(E=1)\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i)\big|E=1\big] + p(E=0)\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i)\big|E=0\big]. \tag{۲}$$

اما برای هر دو دنباله نوعی $(x^n,\hat{x}^n)\in\mathcal{T}(p(x,\hat{x}))$ داریم:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} d(x_i,\hat{x}_i) &= \sum_{x,\hat{x}} \frac{\#i:(x_i,\hat{x}_i)=(x,\hat{x})}{n} d(x,\hat{x}) \\
&= \sum_{x,\hat{x}} \pi(x,\hat{x}|x^n,\hat{x}^n)\cdot d(x,\hat{x}) \\
&\le \sum_{x,\hat{x}} (1+\epsilon)p(x,\hat{x})\cdot d(x,\hat{x}) \\
&= (1+\epsilon)\mathbb{E}_{p(x,\hat{x})}(d(X,\hat{X})) \\
&\le (1+\epsilon)\frac{D}{1+\epsilon} \\
&= D.
\end{aligned}$$

زیرا نوعی بودن دنباله های x^n و $\hat{x}^n$ به این معنی است که تعداد تکرارهای سمبل های $x,\hat{x}$ با مقدار تابع توزیع مشترک احتمال $p(x,\hat{x})$ متناسب است:

$$|\pi(x,\hat{x}|x^n,\hat{x}^n) - p(x,\hat{x})| \le \epsilon p(x,\hat{x}), \qquad \forall x,\hat{x}.$$

بنابراین در صورتی که در معادله (۲) حالت $E=1$ باشد داریم:

$$\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i)\big|E=1\big] \le D.$$

در صورتی که $E=0$ باشد چیزی راجع به میزان اعوجاج نمیتوانیم بگوییم، اما میدانیم که $p(E=0)$ کوچک است. اگر فرض کنیم که $d(x,\hat{x}) < D_{\max}, \quad \forall x,\hat{x}$. داریم:

$$\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i)\big|E=0\big] \le D_{\max}.$$

و در نتیجه از معادله (۲) میتوان نتیجه گرفت که

$$\begin{aligned}
\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i)\big] &\le p(E=1)D + (1-p(E=1))D_{\max} \\
&\le (1-\epsilon)D + \epsilon D_{\max}
\end{aligned}$$

که میتوان آن را با کوچک کردن ϵ به میزان دلخواه به D نزدیک کرد.

۲.۲.۱ قسمت وارون

اگر $R_I(D)$ را تعریف کنیم

$$R_I(D) = \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X})$$

باید نشان دهیم که

$$R(D) \geq R_I(D).$$

یک کد دلخواه به طول n در نظر بگیرید که به اعوجاج $D+\epsilon$ رسیده است. از آنجایی که سمت چپ نامساوی بالا $R(D)$ است، نقطه شروع وارون باید نرخ ارسال کد یعنی R باشد. اگر پیام ارسالی را با M نشان دهیم، داریم:

$$nR = H(M) \geq \cdots \geq n \cdot \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}).$$

که هدف اصلی ما پر کردن نقطه چین ها است. اگر به شکل n حرفی کرانی که میخواهیم ثابت کنیم توجه کنیم میبینیم که بشکل زیر است:

$$\min_{p(\hat{x}^n|x^n):\mathbb{E}[d_n(X^n,\hat{X}^n)]\leq D} I(X^n;\hat{X}^n),$$

که در آن

$$d_n(X^n,\hat{X}^n) = \frac{1}{n}\sum_{i=1}^{n} d(X_i,\hat{X}_i).$$

چون ما یک کد خاص با طول n داریم که در آن یک دنباله X^n توسط فرستنده مشاهده میشود و یک دنباله $\hat{X}^n$ در گیرنده ساخته میشود که شرط $d_n(X^n,\hat{X}^n)$ را برآورده میکند، پس میتوانیم حدس بزنیم که در اثبات از جمله $I(X^n;\hat{X}^n)$ مربوط به همین کد عبور خواهیم کرد (روش سوم اثبات وارون نقطه به نقطه را بیاد آورید):

$$nR = H(M) \geq \cdots \geq I(X^n;\hat{X}^n) \geq \cdots \geq n \cdot \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}).$$

حال باید شروع به پر کردن نقطه چین ها بکنیم. ما زنجیره مارکف $X^n - M - \hat{X}^n$ را داریم که نتیجه میدهد

$$I(X^n;\hat{X}^n|M) = 0$$

از تمرین ۱ میدانیم که اگر در قسمت شرط متغیری را حذف کنیم، اطلاعات متقابل حداکثر به اندازه آنتروپی متغیری که روی آن مشروط کرده ایم تغییر میکند. یعنی:

$$\left|I(X^n;\hat{X}^n|M) - I(X^n;\hat{X}^n)\right| \leq H(M).$$

که با استفاده از $I(X^n;\hat{X}^n|M) = 0$ نتیجه میدهد

$$I(X^n;\hat{X}^n) \leq H(M).$$

پس بخشی از نقطه چین ها پر شده است. باقی میماند که نشان دهیم:

$$I(X^n;\hat{X}^n) \geq \cdots \geq n \cdot \min_{p(\hat{x}|x):\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}).$$

با استدلالی شبیه آنچه در وارون کدگذاری کانال داشتیم، اما با چند تفاوت بارز داریم:

$$\begin{aligned}
I(X^n;\hat{X}^n) &= H(X^n) - H(X^n|\hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i) - H(X^n|\hat{X}^n) && (۳)\\
&= \sum_{i=1}^{n} H(X_i) - H(X_i|\hat{X}^n X_{1:i-1}) \\
&\geq \sum_{i=1}^{n} H(X_i) - H(X_i|\hat{X}_i) \\
&= \sum_{i=1}^{n} I(X_i;\hat{X}_i) && (۴)
\end{aligned}$$

که در معادله (۳) از $i.i.d.$ بودن منبع استفاده کردیم. حال مشابه آن کاری که در کدگذاری کانال با شرط هزینه متوسط داشتیم را انجام میدهیم. اگر تعریف کنیم

$$D_i =: \mathbb{E}[d(X_i,\hat{X}_i)]$$

آنوقت طبق تعریف $R_I(D)$ که شامل یک مینیمم گیری روی همه توزیع های با یک شرط متوسط است، داریم

$$I(X_i;\hat{X}_i) \geq R_I(D_i)$$

پس

$$\sum_{i=1}^{n} I(X_i;\hat{X}_i) \geq \sum_{i=1}^{n} R_I(D_i). \qquad (۵)$$

حال اگر بتوانیم نشان دهیم که $R_I(D)$ محدب است خواهیم داشت.

$$\sum_{i=1}^{n} R_I(D_i) \geq nR_I(\frac{1}{n}\sum D_i). \qquad (۶)$$

اما از آنجایی که کد به اعوجاج $D+\epsilon$ رسیده است داریم:

$$\frac{1}{n}\sum_i D_i = \frac{1}{n}\sum_i \mathbb{E}[d(X_i,\hat{X}_i)] = \mathbb{E}[d_n(X^n,\hat{X}^n)] \leq D$$

پس اگر بتوانیم ثابت کنیم که $R_I(D)$ تابعی غیرصعودی است میتوانیم بنویسیم

$$R_I(\frac{1}{n}\sum D_i) \geq R_I(D). \qquad (۷)$$

معادلات (۴) و (۵) و (۶) و (۷) نتیجه میدهند که

$$I(X^n; \hat{X}^n) \geq R_I(D)$$

که همان چیزی است که میخواستیم آن را ثابت کنیم. اثبات وارون کامل است اگر بتوانیم ثابت کنیم که $R_I(D)$ غیرصعودی و محدب است. اثبات این موضوع شبیه اثبات مربوطه برای کانال های با هزینه ورودی است و به خواننده واگذار میشود.

۳.۱ تابع نرخ - اعوجاج برای متغیرهای برنولی

نیاز به تکمیل