نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

جلسه ششم: لم پوششی

در این جلسه به معرفی لم پوششی میپردازیم. این لم به همراه لم گنجایشی دو ابزار اصلی مرسوم در بخش قابل حصول در نظریه اطلاعات شبکه را تشکیل میدهند. در این بخش ابتدا به مقدمات لم پوششی می پردازیم. سپس شکل کلی آن را بیان میکنیم. در بخش بعد به عنوان اولین کاربرد از لم پوششی، قضیه نرخ اعوجاج را بیان خواهیم کرد.

۱ مقدمات لم پوششی

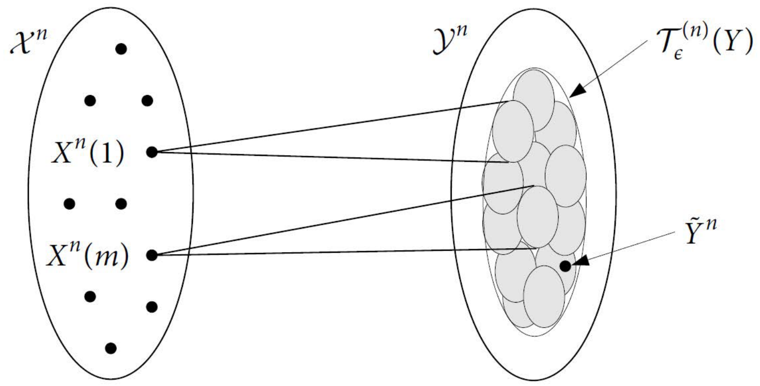
یک توزیع دلخواه $p(x,y)$ در نظر بگیرید. فرض کنید که 2^{nR} دنباله $X^n(1), X^n(2), \cdots, X^n(2^{nR})$ به طول n را به صورت $i.i.d$ از توزیع $p(x)$ بسازیم. در بحث لم گنجایشی دیدیم که در صورتی که اگر $\tilde{y}^n$ یک دنباله نوعی دلخواه از توزیع حاشیه ای $p(y)$ باشد آنوقت

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

حال این سؤال پیش می آید که اگر $R > I(X;Y)$ چه اتفاقی می افتد؟ ثابت خواهیم کرد که در این صورت حد احتمال بجای اینکه به سمت صفر برود، به سمت یک میرود:

$$R > I(X;Y) \Rightarrow \lim_{n\to\infty} Prob[\exists\, 1 \leq m \leq 2^{nR} : (X^n(m), \tilde{y}^n) \in \mathcal{T}_\epsilon^n(p(x,y))] \to 1. \qquad (۱)$$

یعنی هر دنباله نوعی دلخواه $\tilde{y}^n$ با احتمال خیلی بالایی پوشانده میشود. به عبارت دیگر اگر مجموعه نوعی مربوط به هرکدام از کلمات کد را در نظر بگیریم تقریبا تمامی فضای دنباله های نوعی را در سمت $\mathcal{Y}^n$ میپوشاند.



نکته ۱ تفاوت لم پوشاندن را با لم گنجاندن ملاحظه کنید. اگر تعداد نقاط کمتر از $2^{n(I(X;Y)-\epsilon)}$ باشد، دوایر خاکستری رنگ اصلا اشتراکی ندارند و بخش خیلی کوچکی فضا را پوشش می دهند. اما همین که تعداد نقاط از $2^{n(I(X;Y)+\epsilon)}$ بیشتر شد دوایر خاکستری رنگ تمام فضای نوعی را پوشش می دهند. اصطلاحا در اینجا در $R = I(X;Y)$ یک تغییر فاز رخ میدهد. وجه تسمیه این نامگذاری پدیده های فیزیکی مانند یخ زدن آب است. حالت آب در دماهای کمتر از صفر درجه بکلی متفاوت از حالت آب در دماهای بالاتر از صفر درجه میباشد. در اینجا فاز ماده با تغییر دما تغییر میکند. برای برقرار کردن تشبیه، R نقش دما و مقدار $I(X;Y)$ نقش دمای صفر درجه برای آب را بازی میکند.

به عنوان مثال دیگر تابع $f(x,y) = x^y$ را برای $x, y > 0$ در نظر بگیرید. در این صورت

$$\lim_{y\to\infty} f(x,y) = 0, \qquad \forall x < 1,$$

و

$$\lim_{y\to\infty} f(x,y) = \infty, \qquad \forall x > 1.$$

میبینیم که در $x = 1$ یک تغییر فاز رخ میدهد.

اما ماهیت تغییر فازی که در لم گنجایشی و پوششی رخ میدهد در تمرین زیر آورده شده است و بعدا به آن باز خواهیم گشت.

تمرین ۲ فرض کنید که $a > 0$ یک ثابت دلخواه باشد. قرار دهید

$$g(n,x) = (1 - 2^{-na})^{2^{nx}}.$$

ثابت کنید که

$$\lim_{n\to\infty} g(n,x)$$

یک تغییر فاز در $x = a$ دارد:

$$\lim_{n\to\infty} g(n,x) = 0, \qquad \forall x > a,$$
$$\lim_{n\to\infty} g(n,x) = 1, \qquad \forall x < a.$$

راهنمایی: میتوانید از نامساوی $(1-x)^k \le e^{-kx}$ استفاده کنید.

اما جهت اثبات رابطه (؟؟) واقعه E_m را مشابه اثبات لم گنجایشی اینگونه تعریف میکنیم:

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

توجه کنید که هرکدام از $X^n(m)$ ها با احتمال تقریبا $2^{-nI(X;Y)}$ با دنباله y^n مشترکا نوعی میشوند.

$$p(E_m) \approx 2^{-nI(X;Y)}.$$

میخواهیم ثابت کنیم که $p(\bigcup_m E_m) \sim 1$. اما از آنجایی که یک کران پایین روی احتمال اجتماع میخواهیم، دیگر نمیتوانیم از کران اجتماع استفاده کنیم. تعداد کل کلمات کد برابر 2^{nR} است. فرض کنید که دنباله های کد بصورت مستقل از هم و $i.i.d.$ تولید شوند. هر کلمه کد با احتمال $2^{-nI(X;Y)}$ با $\tilde{y}^n$ نوعی میشود. مثل این است که برای هر کلمه کد یک سکه می اندازیم که با احتمال $2^{-nI(X;Y)}$ رو می آید (رو آمدن سکه یعنی نوعی شدن با $\tilde{y}^n$). آنوقت تعداد کلمات کدی که با $\tilde{y}^n$ مشترکا نوعی میشوند یک توزیع دو جمله ای با پارامتر

$$(N = 2^{nR}, p = 2^{-nI(X;Y)})$$

دارد. پس احتمال اینکه حداقل یکی با $\tilde{y}^n$ نوعی شود برابر است با

$$1 - \binom{N}{0} p^0 (1-p)^N = 1 - (1 - 2^{-nI(X;Y)})^{2^{nR}}.$$

باید نشان دهیم که اگر $R > I(X;Y)$ آنوقت $1 - (1 - 2^{-nI(X;Y)})^{2^{nR}} \to 1$ وقتی $n \to \infty$، و یا معادلا

$$(1 - 2^{-nI(X;Y)})^{2^{nR}} \to 0, \qquad as \quad n \to \infty.$$

که با استفاده از تمرین ؟؟ ثابت میشود.

نکته ۳ متوسط تعداد کلمات کدی که با $\tilde{y}^n$ نوعی میشود برابر است با متوسط توزیع دو جمله ای

$$Np = 2^{nR} 2^{-nI(X;Y)} = 2^{n(R - I(X;Y))}.$$

پس وقتی $n \to \infty$ نه تنها یک کلمه کد، بلکه بطور متوسط بینهایت کلمه کد با $\tilde{y}^n$ مشترکا نوعی میشوند.

نکته ۴ احتمال مشترکا نوعی بودن $\tilde{y}^n$ با یک کلمه کد دقیقا $2^{-nI(X;Y)}$ نیست، بلکه تقریبا $2^{-nI(X;Y)}$ است. بصورت دقیق تر این احتمال در نامساوی زیر صدق میکند. اما در بحث شهودی بالا جهت ساده سازی فرض کردیم که احتمال دقیقا $2^{-nI(X;Y)}$ است:

$$2^{-n(I(X;Y)+\delta)} \le p\big((X^n(m), \tilde{y}^n) \in \mathcal{T}_\epsilon(p(x,y))\big) \le 2^{-n(I(X;Y)-\delta)}.$$

نکته ۵ واضح است که اگر $R > I(X;Y)$ باشد، کتاب کد خوبی در مساله کدگذاری کانال نخواهیم داشت. خواهیم دید که کاربرد لم پوششی در مسائل کدگذاری منبع است، در حالی که کاربرد لم گنجایشی در مسائل کدگذاری کانال است.

۱.۱ تعمیم به کدهای ساختار یافته

تعمیم لم پوششی به حالتی که $\tilde{Y}^n$ و $\tilde{U}^n$ دارای توزیعی دلخواه $p(\tilde{u}^n, \tilde{y}^n)$ باشند، را در نظر بگیرید. مشاهده کنید که اگر $(\tilde{U}^n, \tilde{Y}^n)$ مشترکا نوعی نشوند، احتمال اینکه $(\tilde{U}^n, \tilde{Y}^n, X^n(m))$ مشترکا نوعی شوند صفر است (مستقل از اینکه R را چقدر بزرگ کنیم). پس باید یک فرض اضافی در مورد توزیع $p(\tilde{u}^n, \tilde{y}^n)$ گذاشت و آن اینکه

$$\lim_{n\to\infty} p\big((\tilde{U}^n, \tilde{Y}^n) \in \mathcal{T}(p(u,y))\big) = 1. \qquad (۲)$$

مثال ۶ برای برقراری شرط (؟؟) دنباله های $(\tilde{U}^n, \tilde{Y}^n)$ مثلا میتوانند یک دنباله خاص $(\tilde{u}^n, \tilde{y}^n) \in \mathcal{T}(p(u,y))$ را با احتمال یک بگیرند؛ یا اینکه بصورت $i.i.d.$ از توزیع $p(u,y)$ تولید شوند؛ و یا اینکه توزیعی دلخواه (نه لزوما یکنواخت) روی زیرمجموعه ای از دنباله های $\mathcal{T}(p(u,y))$ داشته باشند.

از آنجایی که امکان استفاده از کران اجتماع در لم پوششی وجود ندارد، امکان تعمیم گسترده لم پوششی همانند لم گنجایشی وجود ندارد. بصورت خاص دیگر نمیتوانیم خود را محدود به توزیع حاشیه ای $(X^n(m), \tilde{Y}^n)$ بکنیم. بنابراین باید فرض کرد که دنباله های

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

بصورت مستقل از یکدیگر با گذراندن دنباله $\tilde{U}^n$ از کانال $p_{X|U}$ ساخته میشوند.

مثال ۷ امکان تعمیم مستقیم لم پوششی به زمانی که کلمات کد $X^n(m)$ بصورت دلخواه وابسته به هم تولید شوند وجود ندارد (حداقل در آن شکلی که در مورد لم گنجایشی داشتیم). به عنوان یک مثال نقض فرض کنید که دنباله $X^n(1)$ 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 توزیع $i.i.d.$ $p(y)$ انتخاب کنیم. در این صورت توزیع مشترک $X^n(m)$ و $\tilde{Y}^n$ برابر $\prod_{i=1}^n p_X(x_i)p_Y(\tilde{y}_i)$ است. اما مستقل از اینکه R را چقدر بزرگ کنیم

$$\begin{aligned} Prob[\exists \;\; 1 \leq m \leq 2^{nR} : (X^n(m), \tilde{Y}^n) \in \mathcal{T}_\epsilon^n(p(x,y))] &= Prob[(X^n(1), \tilde{Y}^n) \in \mathcal{T}_\epsilon^n(p(x,y))] \\ &= 2^{-nI(X;Y)} \to 0 \neq 1. \end{aligned}$$

۲ لم پوششی

قضیه ۸ 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 توزیعی 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 اما دارای این خاصیت است که

$$\lim_{n\to\infty} p\big((\tilde{U}^n, \tilde{Y}^n) \in \mathcal{T}(p(u,y))\big) = 1.$$

همچنین فرض کنید که 2^{nR} دنباله های

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

بصورت مستقل با گذراندن دنباله $\tilde{U}^n$ از کانال $p_{X|U}$ ساخته باشیم. یعنی

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

و

$$p(x^n(1), x^n(2), \cdots, x^n(2^{nR})|\tilde{u}^n) = \prod_{m=1}^{2^{nR}} p(x^n(m)|\tilde{u}^n).$$

بعلاوه زنجیره مارکف $X^n(1), X^n(2), \cdots, X^n(2^{nR}) - \tilde{U}^n - \tilde{Y}^n$ برقرار باشد. آنگاه اگر $R > I(X;Y|U)$ داریم:

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

اثبات: واقعه E_m را مشابه اثبات لم گنجایشی اینگونه تعریف میکنیم:

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

میخواهیم ثابت کنیم که

$$\lim_{n\to\infty} p(\bigcup_m E_m) = 1.$$

چون توزیع $p(\tilde{u}^n, \tilde{y}^n)$ دلخواه است به آن مشروط میکنیم:

$$\begin{aligned} p(\bigcup_m E_m) &= \sum_{\tilde{u}^n, \tilde{y}^n} p(\tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n) p(\bigcup_m E_m | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n) \\ &= \sum_{(\tilde{u}^n, \tilde{y}^n) \in \mathcal{T}_\epsilon} p\big(\tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) p\big(\bigcup_m E_m | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) \qquad (۳) \end{aligned}$$

حال جمله درون مجموع را تحلیل میکنیم:

$$\begin{aligned} p\big(\bigcup_m E_m | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) &= 1 - p\big(\bigcap_m E_m^c | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) \qquad (۴) \\ &= 1 - \prod_m p\big(E_m^c | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big) \qquad (۵) \\ &= 1 - \Big(p\big(E_1^c | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big)\Big)^{2^{nR}} \qquad (۶) \\ &= 1 - \Big(1 - p\big(E_1 | \tilde{U}^n = \tilde{u}^n, \tilde{Y}^n = \tilde{y}^n\big)\Big)^{2^{nR}}, \qquad (۷) \end{aligned}$$

که در آن رابطه (؟؟) را با استفاده از قانون دمورگان نوشته ایم؛ رابطه (؟؟) را با استفاده از استقلال کلمات کد بشرط $\tilde{U}^n$ نوشته ایم؛ و نهایتا رابطه (؟؟) را با استفاده از تقارن نوشته ایم. اما

$$\begin{aligned}
p\big(E_1|\tilde{U}^n=\tilde{u}^n,\tilde{Y}^n=\tilde{y}^n\big) &= p\big((X^n(1),\tilde{u}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^n(p(x,u,y))|\tilde{U}^n=\tilde{u}^n,\tilde{Y}^n=\tilde{y}^n\big)\\
&= p\big((X^n(1),\tilde{u}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^n(p(x,u,y))|\tilde{U}^n=\tilde{u}^n\big) &(۸)\\
&= \sum_{x^n} p(x^n|\tilde{U}^n=\tilde{u}^n)\times\\
&\quad p\big((X^n(1),\tilde{u}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^n(p(x,u,y))|\tilde{U}^n=\tilde{u}^n,X^n(1)=x^n\big)\\
&= \sum_{x^n:(x^n,\tilde{u}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^n(p(x,u,y))} p(x^n|\tilde{U}^n=\tilde{u}^n) &(۹)\\
&\geq \sum_{x^n:(x^n,\tilde{u}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^n(p(x,u,y))} 2^{-n(H(X|U)+\delta)} &(۱۰)\\
&= 2^{-n(H(X|U)+\delta)}|\{x^n:(x^n,\tilde{u}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon^n(p(x,u,y))\}|\\
&\geq 2^{-n(H(X|U)+\delta)}2^{n(H(X|UY)-\delta)} &(۱۱)\\
&= 2^{-n(I(X;Y|U)+2\delta)},
\end{aligned}$$

که در آن رابطه (؟؟) را با استفاده از زنجیره مارکف $X^n(1),X^n(2),\cdots,X^n(2^{nR})-\tilde{U}^n-\tilde{Y}^n$ نوشتیم؛ رابطه (؟؟) را با استفاده از این نکته که احتمال مشروط یا صفر و یا یک است، بسته به اینکه $(x^n,\tilde{u}^n,\tilde{y}^n)$ نوعی باشد نوشتیم؛ رابطه (؟؟) را با استفاده از این نکته که $(x^n,\tilde{u}^n)$ مشترکا نوعی هستند و با استفاده از کران پایین روی احتمال دنباله های نوعی شرطی نوشتیم؛ و نهایتا رابطه (؟؟) را با استفاده از کران پایین روی تعداد دنباله های شرطی نوعی نوشتیم.

با جایگذاری در رابطه (؟؟) داریم:

$$p\big(\bigcup_m E_m|\tilde{U}^n=\tilde{u}^n,\tilde{Y}^n=\tilde{y}^n\big)\geq 1-\left(1-2^{-n(I(X;Y|U)+2\delta)}\right)^{2^{nR}},$$

با جایگذاری در رابطه (؟؟) داریم:

$$\begin{aligned}
p(\bigcup_m E_m) &\geq \sum_{(\tilde{u}^n,\tilde{y}^n)\in\mathcal{T}_\epsilon} p\big(\tilde{U}^n=\tilde{u}^n,\tilde{Y}^n=\tilde{y}^n\big)\left(1-\big(1-2^{-n(I(X;Y|U)+2\delta)}\big)^{2^{nR}}\right)\\
&= p\big((\tilde{U}^n,\tilde{Y}^n)\in\mathcal{T}_\epsilon\big)\left(1-\big(1-2^{-n(I(X;Y|U)+2\delta)}\big)^{2^{nR}}\right)
\end{aligned}$$

با گرفتن حد وقتی n به سمت بینهایت میرود، هر دو جمله در عبارت بالا به سمت یک میروند (با توجه به تمرین ؟؟ و اینکه $R>I(X;Y|U)$) و به نتیجه دلخواه میرسیم. اثبات کامل است.

□

در بیان کتاب الجمال و کیم از این قضیه بجای $(\tilde{U},X,\tilde{Y})$ از $(U,\hat{X},X)$ استفاده شده است؛ یعنی بجای $\tilde{Y}$، متغیر X و بجای X متغیر $\hat{X}$ استفاده شده است. دلیل این موضوع نحوه کاربرد آن این قضیه میباشد؛ با این نحوه نمادگذاری

منطبق کردن آن با کاربرد آسان تر میشود. کاربرد لم پوششی در مسائل کدگذاری منبع است، در حالی که کاربرد لم گنجایشی در مسائل کدگذاری کانال است. در بالا ما ابتدا از همان نمادگذاری قضیه گنجایشی استفاده کردیم تا ارتباط میان این دو مشخص تر باشد. اما در زیر صورت قضیه گفته شده با این تغییر متغیر را آورده ایم. در آینده به قضیه زیر به عنوان لم پوششی مراجعه خواهیم کرد:

قضیه ۹ توزیع $p(u,x,\hat{x})$ روی متغیرهای $(U,X,\hat{X})$ را در نظر بگیرید. فرض کنید که دنباله های X^n و U^n دارای توزیعی دلخواه اما دارای این خاصیت است که

$$\lim_{n\to\infty} p\big((U^n,X^n)\in\mathcal{T}(p(u,x))\big)=1.$$

همچنین فرض کنید که 2^{nR} دنباله های

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

بصورت مستقل با گذراندن دنباله U^n از کانال $p_{\hat{X}|U}$ ساخته باشیم. یعنی

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

و

$$p(\hat{x}^n(1),\hat{x}^n(2),\cdots,\hat{x}^n(2^{nR})|u^n)=\prod_{m=1}^{2^{nR}} p(\hat{x}^n(m)|u^n).$$

بعلاوه زنجیره مارکف $\hat{X}^n(1),\hat{X}^n(2),\cdots,\hat{X}^n(2^{nR})-U^n-X^n$ برقرار باشد. آنگاه اگر $R>I(X;\hat{X}|U)$ داریم:

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

۳ تابع نرخ اعوجاج

یکی از مهمترین کاربردهای لم پوششی در مساله تابع نرخ اعوجاج[۱] است. اما برای بیان این مساله نیاز به بیان مقدماتی داریم.

بر اساس قضیه شانون می دانیم برای کدکردن بدون اتلاف منبع گسسته و $i.i.d$ حداقل به اندازه ی آنتروپی منبع بیت بر هر سمبل نیاز داریم. حال اگر منبع پیوسته باشد بازسازی دقیق آن $p(X^n\neq\hat{X}^n)\to 0$ مشکل زا است و با هیچ نرخ فشرده سازی متناهی برآورده نمی شود. در واقع آنتروپی منابع پیوسته بینهایت است (باید بینهایت بیت برای مشخص کردن یک عدد حقیقی انتقال داد) و به همین دلیل ما از آنتروپی تفاضلی استفاده میکردیم. پس باید از تقاضا برای بازیابی دقیق منبع عقب نشینی کنیم، و صرفا بخواهیم که منبع را بصورت تقریبی بازسازی کنیم. به عبارت دیگر بجای آنکه دنبال بازسازی دقیق منبع باشیم، به دنبال تخمین آن (یا بازسازی تقریبی آن) خواهیم بود. به این مساله کدگذاری

[۱] Rate distortion function

منبع با اتلاف[۲] یا فشرده سازی با اتلاف[۳] می گویند، در حالی که کدگذاری منبعی که بازسازی دقیق با احتمال بالا را طلب می نمود، با عنوان کدگذاری منبع بدون اتلاف[۴] یا فشرده سازی بدون اتلاف[۵] شناخته میشود.

در حالتی که تنها یک نسخه از منبع داشته باشیم، فشرده سازی با اتلاف (تقریبی) آن معادل چندی سازی[۶] منبع (تبدیل آنالوگ به دیجیتال) میباشد. اما همانند دیگر مسائل نظریه اطلاعات در صورتی که از کدگذاری بلوکی استفاده کنیم و n نسخه از منبع را بصورت همزمان چندی سازی برداری[۷] کنیم میتوانیم به کارایی بالاتری برسیم.

نرخ بهینه در فشرده سازی بدون اتلاف آنتروپی منبع است. اما در فشرده سازی با اتلاف، نرخ بهینه به میزان اعوجاج مورد نیاز و معیار اعوجاج بستگی دارد. هرچه اعوجاج کمتری بخواهیم، باید از نرخ بیشتری استفاده کنیم. پیدا کردن مقدار بهینه نرخ بر حسب میزان اعوجاج به مساله نرخ-اعوجاج[۸] شهرت دارد. این تابع شبیه تابع هزینه-ظرفیت در کدگذاری کانال است.

اما جهت تعریف دقیق مفهوم بازسازی تقریبی یک منبع نیاز به یک متر فاصله داریم. مثلا با متر مجذور فاصله می توانیم تقاضا کنیم که میانگین فاصله ی مربعات درایه های دنباله اصلی و درایه های دنباله بازسازی شده از حد خاصی کمتر باشد.

تعریف ۱۰ یک متر فاصله تابعی است که دو سمبل x و $\hat{x}$ از منبع را گرفته و یک مقدار حقیقی مثبت $d(x,\hat{x})$ به آنها نسبت می دهد که بیانگر میزان فاصله آن دو سمبل از همدیگر میباشد (به این معنی که اگر بجای x، $\hat{x}$ کدگشایی شود چه میزان هزینه متحمل شده ایم):

$$d : \mathcal{X}\times\mathcal{X}\mapsto[0,\infty).$$

تابع فاصله را میتوان به عنوان جدولی با ابعاد $|\mathcal{X}|\times|\mathcal{X}|$ در نظر گرفت که درون آن با اعداد حقیقی نامنفی پر شده است (به این صورت که درایه $(x,\hat{x})$ آن با $d(x,\hat{x})$ پر شده است). توجه کنید که شرط $d(x_1,x_2)=d(x_2,x_1)$ لزوما برقرار نیست. همچنین اگر چه منطقی است که فرض کنیم $d(x,x)=0$ اما چنین شرطی را در حالت کلی روی تابع اعوجاج قرار نمی دهیم.

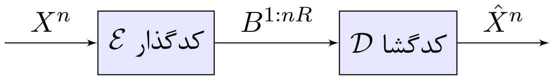
مثال ۱۱ دو متر معروف متر اقلیدسی و متر همینگ هستند. متر اقلیدسی روی $\mathcal{X}=\mathbb{R}$ اینگونه تعریف می شود:

$$d(x,\hat{x})=(x-\hat{x})^2.$$

و متر همینگ روی $\mathcal{X}=\{0,1\}$ اینگونه تعریف می شود: $d(x,\hat{x})=\mathbf{1}[x\neq\hat{x}]$. متر همینگ را میتوان بصورت یک ماتریس بشکل زیر نشان داد:

$$\begin{array}{ccc} 1 & 0 & \\ \hline \multicolumn{1}{|c|}{1} & \multicolumn{1}{c|}{0} & 0 \\ \hline \multicolumn{1}{|c|}{0} & \multicolumn{1}{c|}{1} & 1 \\ \hline \end{array}$$

[۲] Lossy source coding
[۳] Lossy compression
[۴] Lossless source coding
[۵] Lossless compression
[۶] Quantization
[۷] Vector Quantization
[۸] Rate distortion problem



شکل ۱: نمایش شماتیک یک کدگذار منبع برای مسله نرخ-اعوجاج.

که در آن سطرها مربوط به x و ستونها مربوط به $\hat{x}$ هستند.

در حالت منابع گسسته نیز ما ممکن است نخواهیم که بازسازی دقیق انجام دهیم. مثلا ممکن است بخواهیم که پس از کدگذاری و کدگشایی از میان n بیت منبع حداقل نود درصدشان صحیح به مقصد برسند. در این صورت اگر فاصله همینگ را در نظر بگیریم، مقدار متوسط اعوجاج میان سمبل های منبع و سمبل های بازسازی شده اینگونه بدست می آید:

$$\frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i)$$

که میتوانیم تقاضا کنیم که کمتر از ۰.۱ باشد.

۱.۳ تعریف یک کد برای مساله نرخ اعوجاج

تعریف یک کد برای مساله نرخ-اعوجاج شبیه تعریف کد برای مساله کدگذاری منبع است. فرض کنید که منبع X با الفبای $\mathcal{X}$ و یک تابع اعوجاج $d : \mathcal{X} \times \mathcal{X} \mapsto [0, \infty)$ داده شده است. در این صورت یک کد (n, R) برای مساله نرخ اعوجاج متشکل از یک کدگذار برای فشرده‌سازی $x^n \in \mathcal{X}^n$

$$\mathcal{E} : \mathcal{X}^n \mapsto \{1, 2, 3, \cdots, 2^{nR}\}.$$

و یک تابع کدگشا

$$\mathcal{D} : \{1, 2, 3, \cdots, 2^{nR}\} \mapsto \mathcal{X}^n.$$

میباشد (شکل ؟؟ را ببینید). چهارتایی $(n, R, \mathcal{E}, \mathcal{D})$ یک کد را مشخص می‌کند. این کدگذار و کدگشا را میتوان به این شکل نیز بیان کرد. کدگذار به هر دنباله x^n عددی مانند $m \in [1 : 2^{nR}]$ نسبت میدهد، و کدگشا به هر عدد $m \in [1 : 2^{nR}]$ یک دنباله $\hat{x}^n(m)$ نسبت میدهد.

متوسط اعوجاج کد بصورت زیر تعریف میشود:

$$\mathbb{E}\big[\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i)\big] = \sum_{x^n} p(x^n)\big[\frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i)\big]$$

که در آن

$$\hat{X}^n = \mathcal{D} \circ \mathcal{E}(X^n).$$

تعریف ۱۲ یک جفت نرخ-اعوجاج (R, D) قابل حصول خوانده میشود اگر برای هر $\epsilon > 0$ یک دنباله از کدگذارها و کدگشاها با $n \to \infty$ وجود داشته باشد بطوریکه متوسط نرخ اعوجاج کدهای مربوطه از $D + \epsilon$ کمتر باشد.[۹] این تعریف به این معنی است که ما یک فاصله ϵ از D را مجاز میدانیم و تنها در حد $n \to \infty$ به اعوجاج D میرسیم. این شبیه احتمال خطای کدگذاری بدون اتلاف است که بصورت حدی به سمت صفر میرود.

مثال ۱۳ فرض کنید که در تعریف ناحیه نرخ-اعوجاج اضافه کردن ϵ به D را در نظر نمیگرفتیم. اگر X یک توزیع برنولی و d متر همینگ باشد، آنوقت با فرض $D = 0$ و بدون در نظر گرفتن ϵ

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

نتیجه میدهد که $\hat{X}^n = X^n$. یعنی اعوجاج صفر به معنی بازیابی کاملا بدون خطا خواهد بود. بنابراین در نظر گرفتن ϵ در مساله نرخ-اعوجاج همانند در نظر نگرفتن آن در مساله کدگذاری بدون اتلاف است.

تعریف ۱۴ تابع نرخ-اعوجاج $R(D)$ به این صورت تعریف میشود:

$$R(D) = \inf_{(R,D)\ \text{قابل حصول}} R. \tag{۱۲}$$

به این معنی که برای هر اعوجاج D کمترین نرخ ممکن R را می یابیم.

تعریف ۱۵ برای دنباله های x^n و $\hat{x}^n$ عبارت $d_n(x^n, \hat{x}^n)$ اینگونه تعریف میشود

$$d_n(x^n, \hat{x}^n) := \frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i).$$

این نمادگذاری صرفا جهت تسهیل در نوشتن فرمول ها استفاده میشود.

۲.۳ رفتار معمول تابع نرخ-اعوجاج

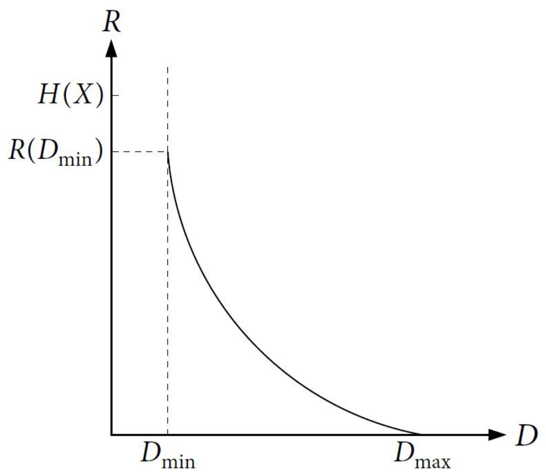
رفتار معمول تابع نرخ-اعوجاج در شکل ؟؟ نشان داده شده است. تابع نرخ اعوجاج $R(D)$ تابعی نزولی بر حسب D است. دلیل این موضوع این است که با افزایش نرخ R دسته بزرگتری از کدها را در نظر میگیریم و امکان رسیدن به اعوجاج های کمتری وجود دارد.

حالت خاص $R \geq H(X)$:

فرض کنید که $R \geq H(X)$. در این صورت فرستنده میتواند منبع را کاملا به گیرنده منتقل کند و در گیرنده با داشتن X^n، دنباله $\hat{X}^n$ را تشکیل میدهد بطوریکه $d_n(X^n, \hat{X}^n)$ کمینه باشد. مثلا اگر

$$d(x, x) = 0, \quad \forall x$$

[۹] این جمله به این معنی است که برای هر N طبیعی کدی با طول n $(n > N)$ وجود دارد بطوریکه نرخ کد R باشد و به اعوجاج $D + \epsilon$ برسیم.



شکل ۲: رفتار معمولی یک تابع نرخ - اعوجاج.

یعنی اعوجاج هر سمبل با خودش صفر باشد، آنوقت گیرنده میتواند دنباله $\hat{X}^n$ را مساوی با X^n قرار دهد و به اعوجاج صفر برسد.

اما مثلا برای یک منبع باینری با متر $d(x,\hat{x}) = \mathbf{1}[x = \hat{x}]$ بهتر است که بیت های X^n را تغییر دهد تا به کمترین اعوجاج برسیم ($\hat{X}_i = 1 - X_i$) . توصیف ماتریسی این متر در جدول زیر آورده شده است:

	1	0
0	1	0
1	0	1

که در آن سطرها مربوط به x و ستونها مربوط به $\hat{x}$ هستند. در متر زیر $\hat{X}^n$ را باید دنباله تمام صفر انتخاب کرد:

	1	0
0	2	2.5
1	2	2.3

با انتخاب بهینه $\hat{X}^n$ به اعوجاج متوسط ۲ میرسیم. اعوجاج از این کمتر قابل دسترسی نیست. در متر زیر اصلا مهم نیست که گیرنده چگونه $\hat{X}^n$ را انتخاب کند:

	1	0
0	1	1
1	1	1

در هر حالت به اعوجاج متوسط یک میرسیم و اعوجاج از این کمتر قابل دسترسی نیست. به کمترین اعوجاج ممکن $D_{\min}$ میگوییم که در شکل هم نشان داده شده است.

توجه کنید که در دو جدول قبلی انتقال اطلاعات با نرخ $R = H(X)$ هم نیاز نیست چون گیرنده از قبل میتواند دنباله خروجی اش را مشخص کند. پس میتوان به اعوجاج حداقلی با نرخ $R = 0$ هم رسید. طبق تعریف $R(D)$، کمترین مقدار R که ما را به اعوجاج $D_{\min}$ میرساند برابر $R(D_{\min})$ بوده و نامساوی زیر را داریم:

$$R(D_{\min}) \leq H(X).$$

تمرین ۱۶ فرض کنید که X مقادیر $\{0,1,2\}$ را با احتمال مساوی میگیرد. برای متر جدول زیر $D_{\min}$ و $R(D_{\min})$ را مشخص کنید. $R(D_{\min})$ را با $H(X)$ مقایسه کنید.

	2	1	0
0	1	0	0
1	1	0	0
2	0	1	1

ب) حال مساله را برای جدول زیر حل کنید:

	2	1	0
0	0	1	1
1	0	0	1
2	1	1	0

حالت خاص $R=0$:

فرض کنید که $R=0$. در این صورت کمترین اعوجاجی که قابل حصول است، محل تقاطع نمودار نرخ-اعوجاج با محور D ها است که در شکل ؟؟ با $D_{\max}$ نشان داده شده است.

مثال ۱۷ فرض کنید که $\mathcal{X}=\mathbb{R}$ و d متر اقلیدسی باشد:

$$d(x,\hat{x})=(x-\hat{x})^2.$$

در این صورت اگر $R=0$ باشد، بهترین تخمین ما از X که $\mathbb{E}\big((X-\hat{X})^2\big)$ را کمینه میکند همان $\mathbb{E}[X]$ میباشد.

$$\hat{X}_i=\mathbb{E}[X].$$

در این صورت متوسط خطای ما واریانس X خواهد بود:

$$D_{\max}=\mathsf{Var}(X).$$

مثلا اگر $X\sim\mathcal{N}(0,1)$ باشد، بهترین تخمین از $\hat{X}^n$ دنباله ای متشکل از n صفر خواهد بود. توجه کنید که این حدس که $\hat{X}$ را در گیرنده بصورت مستقل از X از $\mathcal{N}(0,1)$ تولید کنیم بهینه نیست زیرا متوسط اعوجاج برابر خواهد بود با

$$\mathbb{E}\big((X-\hat{X})^2\big)=\mathsf{Var}(X-\hat{X})=\mathsf{Var}(X)+\mathsf{Var}(\hat{X})=2>\mathsf{Var}(X)=1.$$

تمرین ۱۸ فرض کنید که X توزیع برنولی با پارامتر p را داشته باشد. فرض کنید که متر ما متر همینگ است. در این صورت اگر $R=0$ باشد بهترین تخمین ما از X^n چه خواهد بود؟ نشان دهید که در این حالت

$$D_{\max}=\min(p,1-p).$$

همچنین ثابت کنید که

$$D_{\min}=0, R(D_{\min})=H(X).$$