نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

جلسه ۱

مرور بر مفاهیم اصلی نظریه اطلاعات نقطه به نقطه

در بخش اول درس و پیش از ورود به مباحث مربوط به نظریه اطلاعات شبکه مروری بر مفاهیم اصلی نظریه اطلاعات نقطه به نقطه خواهیم داشت. مرجع مربوط به این بخش فصل ۳ کتاب درسی میباشد.

۱ اهمیت کدگذاری بلوکی

کدگذاری کانال و کدگذاری منبع دو مساله عمده در نظریه اطلاعات میباشند. هدف از این مسائل به ترتیب انتقال اطلاعات روی یک کانال نویزی $p(y|x)$ و فشرده سازی یک منبع اطلاعات X میباشد. جواب به مساله کدگذاری کانال به ساختارهای ترکیبیاتی کانال $p(y|x)$ بستگی خواهد داشت و به نظر مساله مشکلی میباشد. اما در صورتی که نسخه های زیادی از کانال را در دست داشت باشیم و بخواهیم انتقال اطلاعات را روی آن انجام دهیم، میتوان از قانون اعداد بزرگ بهره جست و با تعریف مفاهیمی مانند نوعی بودن به جوابی با شکل ساده برای این مسائل دست یافت.

همچنین وجود نسخه های زیادی از منبع میتواند در کاهش حدی احتمال خطا با افزایش طول کد مفید باشد. منبع برنولی X با توزیع احتمالی در نظر بگیرید. X نتیجه پرتاب یک سکه است.

$$X = \begin{cases} 0, & 0.999 \quad \text{با احتمال} \\ 1, & 0.001 \quad \text{با احتمال} \end{cases}$$

فرض کنید که آذر می‌خواهد نتیجه 1000 پرتاب مستقل این سکه را برای بابک ارسال کند.

آذر بابک

یک راه این است که آذر مستقیما نتیجه هر پرتاب را برای بابک بفرستد که در این صورت 1000 بیت باید ارسال کند. اما راه دیگر این است که از این نکته که احتمال پشت آمدن سکه خیلی کم است استفاده کند. انتظار داریم که به طور متوسط از هر 1000 پرتاب تنها یکی پشت و بقیه رو باشند. بعد از 1000 بار آزمایش این منبع 1001 حالت مختلف برای دنباله نتایج پرتاب قابل تصور است:

$$
\begin{array}{rc}
(1) & 1000\cdots0000 \\
(2) & 0100\cdots0000 \\
(3) & 0010\cdots0000 \\
 & \vdots \\
(1000) & 0000\cdots0001 \\
(1001) & \text{هیچکدام}
\end{array}
$$

حال آذر می‌تواند حالات بالا را شماره‌گذاری کرده و به جای ارسال دنباله‌ی مشاهدات، شماره حالت اتفاق افتاده را (در مبنای دو) ارسال کند. برای انجام این کار $\log_2 1001 \approx 10$ یا تقریبا 10 بیت برای ارسال نیاز دارد. می‌بینیم که از ارسال 1000 بیت به ارسال 10 بیت رسیده‌ایم. تنها هزینه‌ای که آذر می‌پردازد این است که زمانی که آذر "هیچکدام" را ارسال کند، بابک نمی‌تواند دنباله مشاهدات آذر را بازسازی کند. اما اگر احتمال وقوع دنباله‌ای که متناظر با گزینه "هیچکدام" باشد کم باشد، احتمال خطای بابک کم خواهد بود. این مثال مفید بودن کدگذاری بلوکی را نشان می‌دهد.

۲ دنباله‌های نوعی و مجموعه نوعی

جهت استفاده از کدگذاری بلوکی در سناریوهای کلی لازم است رفتار پرتاب‌های مکرر یک سکه (و در حالت کلی‌تر نمونه‌های i.i.d. یک متغیر تصادفی) را بررسی کنیم. یکی از مفاهیم اصلی که در پرتاب‌های مکرر بروز می‌کند مفهوم نوعی بودن است.

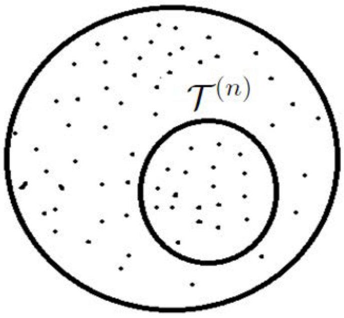
۱.۲ حالت دودویی: بحث شهودی

برای بیان مفهوم نوعی بودن با آزمایش پرتاب سکه شروع می‌کنیم که می‌توان آن را با یک متغیر تصادفی برنولی مدل کرد. فرض کنید متغیر برنولی‌ای داریم که توزیع احتمال آن به شکل زیر است.

$$X = \begin{cases} 0, & \text{با احتمال } 1-p \\ 1, & \text{با احتمال } p \end{cases}$$

از پرتاب n نسخه مستقل این سکه دنباله‌ای n بیتی مانند 001100...0 بوجود می‌آید. تعداد دنباله‌های n بیتی از صفر و یک برابر 2^n است، اما با توجه به توزیع احتمال انتظار داریم که به طور متوسط np تا از بیت‌ها یک باشند. یعنی انتظار داریم که دنباله‌های خاصی از 2^n دنباله ممکن اتفاق بیفتند. پس به طور شهودی انتظار داریم که دنباله‌هایی از مجموعه زیر را ببینیم:

$$\mathcal{T}^{(n)} = \big\{(x_1, x_2, \cdots, x_n) \in \{0,1\}^n : \#(x_i = 1) = np\big\}.$$



شکل ۱: مجموعه نوعی زیرمجموعه‌ای از تمام دنباله‌های ممکن n تایی از صفر و یک است. اندازه این زیرمجموعه تقریبا برابر $2^{nH(X)}$ است که در مقایسه با اندازه کل مجموعه دنباله ها 2^n خیلی کوچکتر است.

البته دقت کنید انتظاری که داریم این است که نتیجه n پرتاب نزدیک np تا یک دارد و نه دقیقا np تا یک (مثلا ممکن است $np+1$ تا یک داشته باشد). اما در ابتدا از این نکته اغماض می‌کنیم. (در اینجا هدف انتقال شهود بدون ذکر جزئیات است. در بحثی دقیق که در بخش بعد آمده به این نکات توجه می‌کنیم.) اندازه‌ی این مجموعه برابر است با:

$$\left|\mathcal{T}^{(n)}\right| = \binom{n}{np} = \frac{n!}{np!(n-np)!}$$

که با استفاده از تقریب استرلینگ از فاکتوریل ($n! \approx (\frac{n}{e})^n$) می‌توان آن را ساده کرد:

$$\begin{aligned}\left|\mathcal{T}^{(n)}\right| = \frac{n!}{np!(n(1-p))!} &\approx \frac{(\frac{n}{e})^n}{(\frac{np}{e})^{np}(\frac{n(1-p)}{e})^{n(1-p)}}\\ &= \frac{1}{p^{np}(1-p)^{n(1-p)}}\\ &= 2^{-np\log(p)}2^{-n(1-p)\log(1-p)}\\ &= 2^{n[-p\log(p)-(1-p)\log(1-p)]}\\ &:= 2^{nH(X)}\end{aligned}$$

که به عبارت $-p\log(p)-(1-p)\log(1-p)$ آنتروپی متغیر تصادفی گفته و با نماد $H(X)$ نمایش داده می‌شود. پس نتیجه می‌گیریم که با اینکه تعداد دنباله‌های n تایی از صفر و یک 2^n است، اما در عمل (با احتمال زیاد) یکی از $2^{nH(X)}$ دنباله مجموعه $\mathcal{T}^{(n)}$ اتفاق می‌افتد. به مجموعه $\mathcal{T}^{(n)}$، مجموعه دنباله‌های نوعی یا دنباله‌های متعارف گفته می‌شود چون انتظار داریم که معمولا اتفاق بیفتند.

دقت کنید که بحث بالا تعریفی از آنتروپی بدست می‌دهد چرا که یک تعریف از آنتروپی یک منبع، اندازه مجموعه نوعی آن می‌باشد. دقت کنید که برای یک منبع دودویی مثل یک سکه همواره $H(X) \leq 1$. اگر $H(X) < 1$ باشد تعداد اعضای مجموعه نوعی $2^{nH(X)}$ از تعداد اعضای کل مجموعه (یعنی 2^n) خیلی کمتر است. در واقع مجموعه نوعی درصد (از مرتبه نمایی) کوچکی از کل دنباله‌ها است. در حالت حدی

$$\lim_{n\to\infty}\frac{2^{nH(X)}}{2^n} = \lim_{n\to\infty} 2^{-n(1-H(X))} = 0.$$

به علاوه احتمال وقوع هر عضو دنباله نوعی با استفاده از اصل ضرب برابر است با

$$p^{np}(1-p)^{n(1-p)} = 2^{np\log(p)}2^{n(1-p)\log(1-p)} = 2^{-n[-p\log(p)-(1-p)\log(1-p)]} = 2^{-nH(X)}$$

پس مجموعه نوعی $2^{nH(X)}$ دنباله دارد که احتمال وقوع هر کدام $2^{-nH(X)}$ است. یعنی توزیع احتمال روی مجموعه نوعی یکنواخت است.

۱.۱.۲ حالت دودویی: بحث دقیق

حال می‌خواهیم بحث‌های شهودی که تا اینجا انجام دادیم را به صورت ریاضی‌وار دقیق کنیم. پرتاب n نسخه مستقل متغیر تصادفی X را در نظر بگیرید. اگر دنباله پرتاب سکه را با متغیرهای تصادفی $X_1, X_2, \cdots, X_n$ نشان دهیم، داریم:

$$\lim_{n\to\infty} \frac{X_1+X_2+\cdots+X_n}{n} = \mathbb{E}[X] = p, \qquad (\text{با احتمال } 1).$$

توجه کنید که $\frac{X_1+X_2+\cdots+X_n}{n}$ همان تعداد یک‌ها در دنباله پرتاب سکه است. از عبارت بالا نتیجه می‌شود که برای هر $\epsilon > 0, \delta > 0$ دلخواه، N به اندازه کافی بزرگ وجود دارد بطوریکه برای هر $n > N$

$$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n} - p\right| > \delta\right) \le \epsilon. \tag{۱}$$

با جایگذاری $\delta = \epsilon p$ رابطه بالا را به شکل زیر می‌نویسیم:[۱] برای هر $\epsilon > 0$ دلخواه، N به اندازه کافی بزرگ وجود دارد به طوری که برای هر $n > N$

$$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n} - p\right| > p\epsilon\right) \le \epsilon.$$

همچنین اگر بجای درصد یک‌ها به درصد صفرها علاقه‌مند باشیم، کافی است رابطه بالا را به شکل زیر بنویسیم:

$$P\left(\left|\frac{\overline{X}_1+\overline{X}_2+\cdots+\overline{X}_n}{n} - (1-p)\right| > (1-p)\epsilon\right) \le \epsilon. \tag{۲}$$

که در آن $\overline{X}_i = 1 - X_i$.

پیش از ادامه دادن لازم است کمی در مورد نمادگذاری صحبت شود. یک دنباله خاص $(x_1, x_2, \cdots, x_n) \in \{0,1\}^n$ را برای راحتی با $x^n = (x_1, x_2, \cdots, x_n)$ نشان داده و دنباله متغیرهای تصادفی $(X_1, X_2, \cdots, X_n)$ را با X^n نشان می‌دهیم.

برای یک دنباله خاص مانند x^n 0100100100 درصد صفرها را با $\pi(0|x^n)$ و درصد یک‌ها را با $\pi(1|x^n)$ نشان می‌دهیم. (پس مثلا تعداد یک‌ها در x^n برابر است با $\pi(1|x^n)n$.) در حالت کلی وقتی می‌نویسیم $\pi(x|x^n)$ منظور تعداد تکررهای x در دنباله x^n است. برای مثلا اگر $x^n = 01001010$ آنگاه

$$\pi(x|x^n) = \begin{cases} \frac{5}{8}, & x = 0 \\ \frac{3}{8}, & x = 1. \end{cases}$$

[۱] اینکه δ را ϵp انتخاب کرده ایم، چندان مهم نیست و تنها به دلیلی تکنیکی و جهت سهولت انتخاب شده است. توجه کنید که زمانی که $p = 0$ است، قطعا هیچوقت یک در پرتاب ها اتفاق نمی افتد و این انتخاب مشکلی ندارد.

روابط (۱) و (۲) نتیجه می‌دهند که

$$P(|\pi(1|X^n) - p| > p\epsilon) \leq \epsilon.$$
$$P(|\pi(0|X^n) - (1-p)| > (1-p)\epsilon) \leq \epsilon.$$

پس طبق باند مجموع[۲]

$$P\bigg(|\pi(1|X^n) - p| > p\epsilon \;\text{ یا }\; |\pi(0|X^n) - (1-p)| > (1-p)\epsilon\bigg) \leq 2\epsilon.$$

یا

$$P\bigg(|\pi(1|X^n) - p(X=1)| \leq p\epsilon \;\text{ و }\; |\pi(0|X^n) - p(X=0)| \leq (1-p)\epsilon\bigg) \geq 1 - 2\epsilon.$$

یعنی اگر مجموعه دنباله‌هایی را در نظر بگیریم که

$$\mathcal{T}_\epsilon^{(n)} = \{x^n : |\pi(1|x^n) - p(X=1)| \leq p\epsilon, \;\; |\pi(0|x^n) - p(X=0)| \leq (1-p)\epsilon\}.$$

داریم:

$$P_{X^n}(\mathcal{T}_\epsilon^{(n)}) \geq 1 - 2\epsilon.$$

به علاوه با استفاده از استدلال‌هایی شبیه آنچه در بالا آمده می‌توان نشان داد:

$$2^{n(H(X)-\epsilon)} \leq \left|\mathcal{T}_\epsilon^{(n)}\right| \leq 2^{n(H(X)+\epsilon)}.$$

همچنین ثابت می‌شود که دنباله‌های نوعی تقریبا هم احتمال هستند:

$$2^{-n(H(X)+\epsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\epsilon)}, \quad \forall x^n \in \mathcal{T}_\epsilon^{(n)}.$$

می‌دانیم که احتمال کل مجموعه نوعی نزدیک 1 است. پس جمع احتمال دنباله‌های موجود در مجموعه $\mathcal{T}_\epsilon^{(n)}$ روی هم نزدیک 1 خواهد بود.

مثال ۱ اگر $p = \frac{2}{3}$ باشد، آنگاه در پرتاب n بار سکه، محتمل‌ترین دنباله‌ی n بیتی دنباله تمام یک می‌باشد. ولی این دنباله خارج از مجموعه نوعی $\mathcal{T}_\epsilon^{(n)}$ است! احتمال دنباله تمام یک برابر است با $(\frac{2}{3})^n$ است در حالی که احتمال هر عضو مجموعه $\mathcal{T}_\epsilon^{(n)}$ تقریبا برابر است با $(\frac{1}{3})^{\frac{1}{3}n}(\frac{2}{3})^{\frac{2}{3}n}$. این تناقض نیست، چون اگر چه دنباله‌هایی همانند دنباله تمام یک وجود دارند که احتمالشان زیادتر از احتمال دنباله‌های نوعی است، اما تعداد چنین دنباله‌هایی خیلی کم است و جمع احتمال همه آنها روی هم نزدیک صفر می‌شود.

[۲] Union Bound

مثال ۲ در حالت خاص $p = \frac{1}{2}$ ، مجموعه نوعی مجموعه‌ای است که نیمی صفر و نیمی یک داشته باشد. در این حالت با توجه به فرمول آنتروپی دودویی که در بالا بدست آوردیم $H(X) = 1$ و در نتیجه تعداد اعضای مجموعه نوعی برابر $2^n \approx \binom{n}{\frac{n}{2}}$ است. در واقع در بسط

$$2^n = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} \approx \binom{n}{\frac{n}{2}}$$

دیگر جملات در مقابل $\binom{n}{\frac{n}{2}}$ قابل صرف نظر کردن هستند. رابطه $2^n \approx \binom{n}{\frac{n}{2}}$ را بگونه دیگری نیز می‌توان ثابت کرد. وقتی $p = \frac{1}{2}$ احتمال تمامی دنباله‌ها برابر خواهد بود با $(\frac{1}{2})^n$. یعنی، همه دنباله‌ها هم احتمال هستند (چه نوعی و چه غیر نوعی). پس احتمال مجموعه نوعی برابر است با تعداد اعضای آن ضربدر احتمال هر عضو آن: $(\frac{1}{2})^n \times \binom{n}{\frac{n}{2}}$. اما طبق مطالب بالا احتمال کل مجموعه نوعی نزدیک 1 است، پس

$$(\frac{1}{2})^n \times \binom{n}{\frac{n}{2}} \approx 1.$$

راه دیگر دیدن شهودی این رابطه از تقریب گوسی برای توزیع دو جمله‌ای است. از آن جایی که تابع گوسی به شکل e^{-x^2} افت می‌کند، مساحت زیر نمودار از یک نقطه x_0 به بعد در مقایسه با مقدار تابع گوسی در نقطه x_0 خیلی کوچکتر خواهد بود. در مورد توزیع نمایی (که بشکل e^{-x} افت می‌کند) این دو مقدار متناسب هستند:

$$\lim_{x_0\to\infty} \frac{\int_{x_0}^{\infty} e^{-\lambda x} dx}{e^{-\lambda x_0}} = \frac{1}{\lambda} > 0.$$

اما در مورد توزیع گوسی که افت شدیدتر است نسبت این دو به سمت صفر می‌رود:

$$\lim_{x_0\to\infty} \frac{\int_{x_0}^{\infty} e^{-\frac{x^2}{\sigma^2}} dx}{e^{-\frac{x_0^2}{\sigma^2}}} = 0.$$

۳ مجموعه نوعی قوی و ضعیف

دو روش متفاوت برای تعریف دنباله‌های نوعی در متون تئوری اطلاعات وجود دارد که به نوعی بودن قوی و ضعیف[۳] معروف هستند. مجموعه‌های نوعی قوی زمانی که با الفبای گسسته و محدود سر و کار داشته باشیم مفید هستند. برای بحث در مورد الفبای پیوسته معمولا از مفهوم نوعی بودن ضعیف استفاده می‌شود.

در حالت کلی وقتی X دودویی نیست، مجموعه نوعی قوی این گونه تعریف می‌شود:[۴]

$$\mathcal{T}_{\epsilon}^{(n)}(p(x)) = \{x^n : |\pi(x|x^n) - p(x)| \leq \epsilon p(x), \quad \forall x \in \mathcal{X}\}.$$

دقت کنید که مجموعه نوعی تنها به n، ϵ و توزیع احتمال $p(x)$ ربط دارد. اما برخی کتاب‌ها برای سادگی آن را با $\mathcal{T}_{\epsilon}^{(n)}(X)$ نشان می‌دهند که این البته یک نمادگذاری است چون مجموعه نوعی به خود متغیر تصادفی X ربطی ندارد،

[۳] Strong and weak typicality

[۴] اینکه از $\epsilon p(x)$ استفاده کرده ایم، تضمین میدهد که اگر احتمال سمبلی صفر باشد، اصلا (و حتی یک بار هم) این سمبل در دنباله های نوعی ظاهر نمیشود.

و فقط به توزیع آن مربوط است. مشابه حالت دودویی داریم:

$$\lim_{n\to\infty} P_{X^n}(\mathcal{T}_\epsilon^{(n)}) = 1.$$

$$2^{n(H(X)-\epsilon)} \le \left|\mathcal{T}_\epsilon^{(n)}\right| \le 2^{n(H(X)+\epsilon)}.$$

$$2^{-n(H(X)+\epsilon)} \le p(x^n) \le 2^{-n(H(X)-\epsilon)}, \quad \forall x^n \in \mathcal{T}_\epsilon^{(n)}.$$

که در آن

$$H(X) = -\sum_x p(x)\log(p(x)).$$

به این عبارت آنتروپی (شانون) گفته می‌شود.

مجموعه نوعی ضعیف این گونه تعریف می‌شود:

$$\mathcal{T}_\epsilon^{(n)}(p(x)) = \{x^n : |-\frac{1}{n}\log(p(x^n)) - H(X)| \le \epsilon\}.$$

که در آن $\frac{1}{n}\log(p(x^n))$ با استفاده از توزیع $i.i.d.$ مربوط به $p(x)$ داده شده محاسبه میشود:

$$\frac{1}{n}\log(p(x^n)) = \frac{1}{n}\sum_{i=1}^{n}\log(p(x_i))$$

دقت کنید که نوعی بودن ضعیف تنها یک شرط روی دنباله اعمال میکند، در حالی که نوعی بودن قوی به اندازه الفبای $|\mathcal{X}|$ شرط روی دنباله اعمال میکند.

نکته ۳ نوعی بودن قوی نوعی بودن ضعیف را نتیجه میدهد. اما بر عکس آن درست نیست. منبع برنولی با $p = \frac{1}{2}$ را در نظر بگیرید. برای یافتن مجموعه نوعی ضعیف توجه کنید که $H(X) = 1$ و در نتیجه اعضای مجموعه نوعی ضعیف آنهایی هستند که احتمال آنها حدودا $2^{-nH(X)} = 2^{-n}$ باشد. وقتی $p = \frac{1}{2}$ احتمال تمامی دنباله‌ها برابر $(\frac{1}{2})^n$ خواهد بود (همه دنباله‌ها هم احتمال هستند). در نتیجه مجموعه نوعی ضعیف شامل تمامی دنباله ها میباشد، در حالی که مجموعه نوعی قوی مجموعه‌ای است که نیمی صفر و نیمی یک داشته باشد.

۴ مجموعه نوعی برای دو یا چند متغیر

مجموعه‌های نوعی قوی و ضعیف را برای بیش از دو متغیر هم می‌توان تعریف کرد.

۱.۴ نوعی قوی

اگر دو دنباله خاص x^n, y^n داشته باشیم درصد تعدادهای تکرار x و y (با هم) را با نماد $\pi(x,y|x^n,y^n)$ نشان میدهیم. با استفاده از این نمادگذاری مجموعه نوعی قوی این گونه تعریف می‌شود

$$\mathcal{T}_\epsilon^{(n)}(p(x,y)) = \{(x^n,y^n) : \left|\pi(x,y|x^n,y^n) - p(x,y)\right| \le \epsilon p(x,y), \quad \forall x\in\mathcal{X}, y\in\mathcal{Y}\}.$$

اعضای $\mathcal{T}_\epsilon^{(n)}$ مشترکا نوعی خوانده می‌شوند. به طور مشابه

$$\lim_{n\to\infty} P_{X^n,Y^n}(\mathcal{T}_\epsilon^{(n)}(p(x,y))) = 1.$$

$$2^{n(H(X,Y)-\epsilon)} \leq \left|\mathcal{T}_\epsilon^{(n)}(p(x,y))\right| \leq 2^{n(H(X,Y)+\epsilon)}.$$

$$2^{-n(H(X,Y)+\epsilon)} \leq p(x^n,y^n) \leq 2^{-n(H(X,Y)-\epsilon)}, \qquad \forall x^n y^n \in \mathcal{T}_\epsilon^{(n)}(p(x,y)).$$

مثال ۴ فرض کنید منابع برنولی X,Y با توزیع مشترک زیر مفروض باشند:

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

حال اگر یک مشاهده از این منبع مشترک را داشته باشیم به صورت متعارف می‌بایست تعداد $(x=0,y=0)$ ها np_{00} و تعداد $(x=1,y=0)$ ها np_{10} و تعداد $(x=0,y=1)$ ها np_{01} و تعداد $(x=1,y=1)$ ها np_{11} باشد. یک زوج دنباله (x^n,y^n) که دارای این خاصیت است مشترکا نوعی خوانده می‌شود. اگر (x^n,y^n) مشترکا نوعی قوی باشد هرکدام از x^n و y^n نیز (به تنهایی) نوعی قوی است زیرا به عنوان مثال تعداد صفرها در دنباله x^n برابر است با:

$$np_{00} + np_{01} = n(p_{00}+p_{01}) = nP(X=0).$$

۲.۴ نوعی ضعیف

دو دنباله خاص x^n,y^n را مشترکا نوعی ضعیف میگوییم اگر

$$p(x^n,y^n) \approx 2^{-nH(X,Y)}, \qquad p(x^n) \approx 2^{-nH(X)}, \qquad p(y^n) \approx 2^{-nH(Y)}.$$

بصورت دقیق

$$\begin{aligned}\mathcal{T}_\epsilon^{(n)}(p(x,y)) = \Big\{(x^n,y^n) :& \big|\frac{1}{n}\log_2 p(x^n,y^n) - H(X,Y)\big| \leq \epsilon \\ & \big|\frac{1}{n}\log_2 p(x^n) - H(X)\big| \leq \epsilon \\ & \big|\frac{1}{n}\log_2 p(y^n) - H(Y)\big| \leq \epsilon\Big\}.\end{aligned}$$

نکته ۵ بر خلاف نوعی بودن قوی، اعمال شرط $p(x^n,y^n) \approx 2^{-nH(X,Y)}$ روی دو دنباله، شرطهای $p(x^n) \approx 2^{-nH(X)}$ و $p(y^n) \approx 2^{-nH(Y)}$ را بصورت مستقیم نتیجه نمیدهد و نیاز داریم که آنها را نیز در تعریف نوعی بودن بگنجانیم.

۳.۴ نوعی بودن شرطی قوی

نوعی بودن شرطی قوی را با یک مثال بیان می‌کنیم و تعمیم آن در حالت کلی را به خواننده واگذار می‌کنیم. فرض کنید منابع برنولی X, Y با توزیع مشترک زیر مفروض باشند:

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

حال اگر دنباله‌ی نوعی $x^n = 0100110\cdots 0$ داده شده باشد، چند دنباله مانند y^n وجود دارد که با x^n نوعی است؟ فرض کنید جاهایی که x^n صفر و یک است را از هم جدا کنیم

$$\begin{array}{lllllllll} x^n = & 0 & 1 & 0 & 0 & 1 & 1 & 0 \cdots & 0 \\ & 0 & & 0 & 0 & & & 0 \cdots & 0 \\ & & 1 & & & 1 & 1 & \cdots & \end{array}$$

اگر (x^n, y^n) مشترکا نوعی باشد، طبق تعریف تعداد 1-هایی که در y^n متناظر با 0-های دنباله x^n می‌آیند np_{01} است. به همین ترتیب تعداد 1-هایی که در y^n متناظر با 1-های دنباله x^n می‌آیند برابر np_{11} است. پس طبق اصل ضرب تعداد این دنباله‌ها برابر است با:

$$\begin{aligned} \binom{nP(X=0)}{np_{01}}\binom{nP(X=1)}{np_{11}} &= \frac{(nP(X=0))!}{(np_{00})!(np_{01})!}\frac{(nP(X=1))!}{(np_{10})!(np_{11})!} \\ &= \frac{n!}{(np_{00})!(np_{01})!(np_{10})!(np_{11})!}\frac{(nP(X=0))!(nP(X=1))!}{n!} \\ &\approx 2^{nH(X,Y)}\frac{1}{2^{nH(X)}} \\ &:= 2^{nH(Y|X)}. \end{aligned}$$

دقت کنید که رابطه بالا توجیهی برای تعریف $H(Y|X) := H(X,Y) - H(X)$ است.

آنتروپی شرطی: عبارت $H(Y|X)$ به صورت زیر محاسبه می‌شود:

$$\begin{aligned} H(Y|X) &= \sum_{x,y} p(x,y)\log\frac{1}{p(x,y)} - \sum_{x} p(x)\log\frac{1}{p(x)} \\ &= \sum_{x,y} p(x,y)\log\frac{1}{p(x,y)} - \sum_{x,y} p(x,y)\log\frac{1}{p(x)} \\ &= \sum_{x,y} p(x,y)\log\frac{p(x)}{p(x,y)} \\ &= \sum_{x,y} p(x,y)\log\frac{1}{p(y|x)} \end{aligned}$$

که برابر است با

$$\begin{aligned}H(Y|X) &= \sum_{x,y} p(x,y)\log\frac{1}{p(y|x)}\\ &= \sum_{x} p(x)\sum_{y} p(y|x)\log\frac{1}{p(y|x)}\\ &:= \sum_{x} p(x)H(Y|X=x)\end{aligned}$$

که در آن

$$H(Y|X=x) := \sum_{y} p(y|x)\log\frac{1}{p(y|x)}$$

آنتروپی Y مشروط به واقعه $X=x$ است. در اینجا از نمادگذاری مرسوم حروف بزرگ برای متغیرهای تصادفی و از حروف کوچک برای مقادیر آنها استفاده می‌کنیم.

نکته ۶ هیچ گاه نمی‌توان گفت که متغیرهای تصادفی X^n و Y^n نوعی هستند. چنین مفهومی تعریف نشده است: تنها دو دنباله می‌توانند نوعی باشند و نه دو متغیر تصادفی. اما عبارت "احتمال اینکه متغیرهای تصادفی X^n و Y^n نوعی باشند $P\big((X^n,Y^n)\in\mathcal{T}_\epsilon\big)$ " معنی‌دار است. معنی این عبارت این است که احتمال اینکه دنباله تصادفی حاصل از آزمایش X^n و Y^n نوعی از کار در بیایند چقدر است. مقدار این احتمال برابر است با

$$P\big((X^n,Y^n)\in\mathcal{T}_\epsilon^{(n)}\big) = \sum_{(x^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} p(x^n,y^n).$$

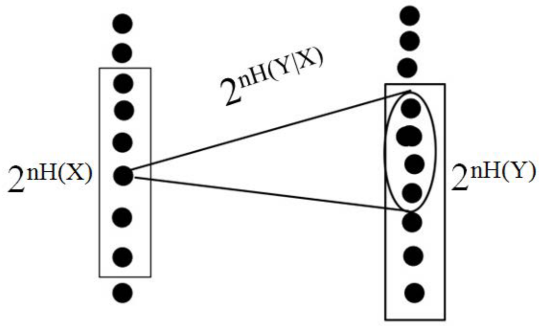
۴.۴ گراف نوعی

گراف نوعی[۵] گرافی دو بخشی است به طوری که در یک بخش دنباله‌های x^n را به عنوان رئوس و در بخش دیگر دنباله‌های y^n را به عنوان رئوس قرار می‌دهیم. دو راس به هم متصل هستند اگر دنباله‌های متناظر آن دو راس مشترکا نوعی باشند. مجموعه نوعی و درجه هر راس در شکل ۲ نشان داده شده است. برای هر دنباله نوعی x^n، تعداد دنباله y^n که با x^n نوعی باشد $2^{nH(Y|X)}$ است.

با توجه به این شکل اگر دنباله‌ی x^n نوعی باشد و دنباله‌ی Y^n بصورت i.i.d. از $p(y)$ (توزیع حاشیه‌ای Y) تولید شود احتمال این که x^n و Y^n مشترکا نوعی شوند برابر است با $2^{-nI(X;Y)}$. دلیل این موضوع این است که Y^n توزیع یکنواخت روی دنباله‌های نوعی‌اش خواهد داشت؛ یعنی توزیع یکنواخت روی $2^{nH(Y)}$ دنباله. از این میان $2^{nH(Y|X)}$ دنباله با x^n نوعی هستند پس

$$P\big((x^n,Y^n)\in\mathcal{T}_\epsilon^{(n)}\big) \approx \frac{2^{nH(Y|X)}}{2^{nH(Y)}} = 2^{-nI(X;Y)}.$$

[۵] Typicality Graph



شکل ۲: گراف نوعی

۵ آنتروپی، نامساوی های آنتروپی و عبارات آنتروپی

دقت کنید که عبارات آنتروپی را میتوان بگونه ای نوشت که تنها بخش مربوط به لگاریتم آنها تفاوت کند:

$$\begin{aligned}
H(X) &= \sum_x p_X(x) \log \frac{1}{p_X(x)} = \sum_{x,y} p_{X,Y}(x,y) \log \frac{1}{p_X(x)} \\
H(Y|X) &= \sum_{x,y} p_{X,Y}(x,y) \log \frac{1}{p_{Y|X}(y|x)} \\
H(X,Y) &= \sum_{x,y} p_{X,Y}(x,y) \log \frac{1}{p_{X,Y}(x,y)} \\
I(X;Y) &= \sum_{x,y} p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}
\end{aligned}$$

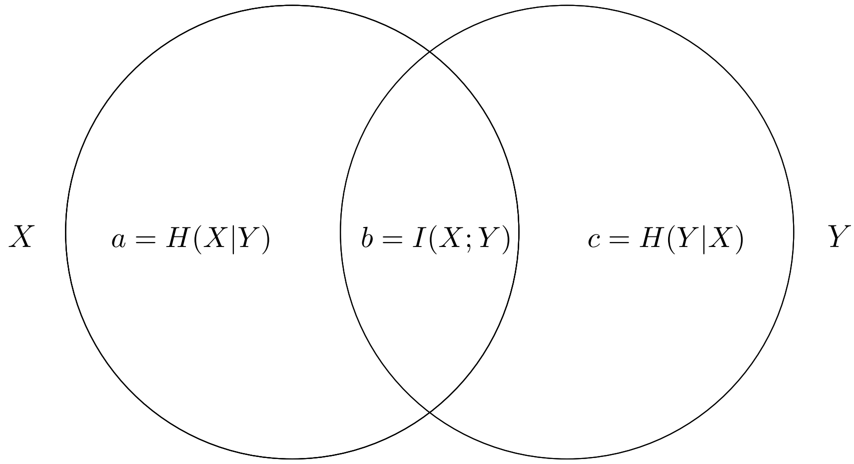
اطلاعات متقابل بین X و Y برابر است با:

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

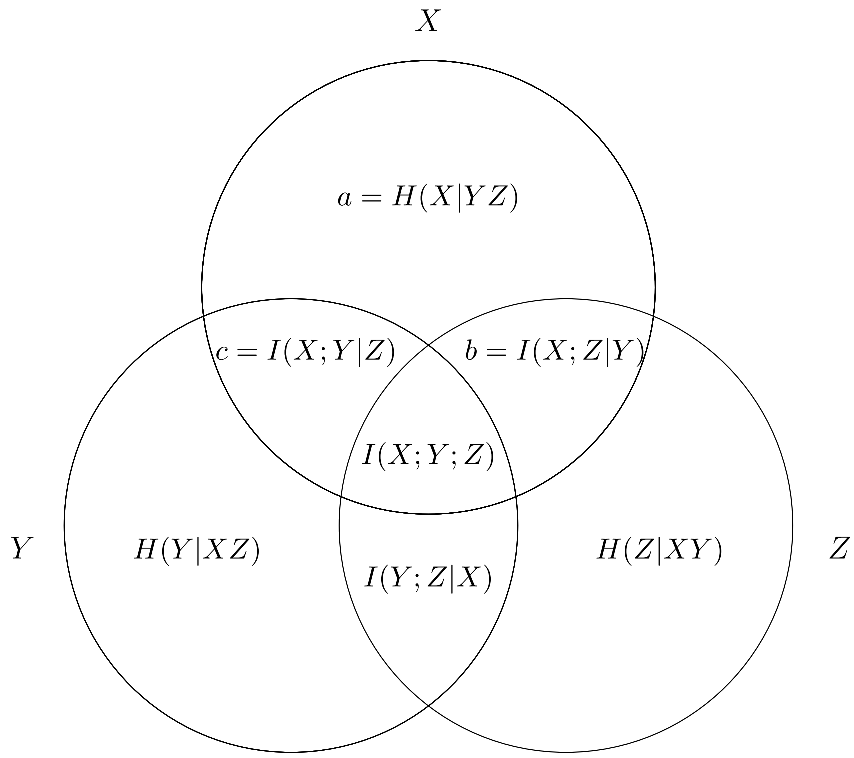
۱.۵ نمودار ون و نامساوی های آنتروپی با حداکثر سه متغیر

نمودار ون برای دو متغیر X و Y در شکل ۱.۵ آمده است. خواهیم دید که استفاده از این نمودار در مسایل مختلف حل را آسان می نماید. نمودار ون برای سه متغیر در شکل ۱.۵ آمده است که به عنوان مثال در آن $I(X;Y|Z) = H(X|Y) - H(X|Y,Z)$. در نمودار ون وقتی مشروط به متغیری میکنیم، به ناحیه خارج از آن متغیر نگاه میکنیم. اما این سؤال پیش می آید که قسمت وسط نمودار ون که با $I(X;Y;Z)$ نشان داده شده متناظر با چه عبارتی است؟ میتوان قسمت وسط را حساب کرد:

$$I(X;Y) - I(X;Y|Z) = I(X;Z) - I(X;Z|Y) = I(Y;Z) - I(Y;Z|X).$$



شکل ۳: نمودار ون برای دو متغیر کلاسیک. تمامی سه قسمتی که نشان داده شده نامنفی هستند.



شکل ۴: نمودار ون برای سه متغیر. تنها ناحیه وسط $I(X;Y;Z)$ میتواند منفی شود.

متاسفانه بر خلاف قسمت های دیگر این قسمت وسط می تواند منفی بشود! مثلا فرض کنید که X و Y دو متغیر برنولی مستقل یکنواخت هستند و Z برابر XOR آنها است. در این صورت

$$I(X;Y)=0, I(X;Y|Z)=H(X|Z)-H(X|Y,Z)=1-0=1$$

پس $I(X;Y)-I(X;Y|Z)=-1$. پس تصور اطلاعات به عنوان چیزی شبیه یک مجموعه، تصوری غلط است. اما از نمودار ون میتوان جهت اثبات نامساوی های آنتروپی با سه متغیر استفاده کرد، اگر فراموش نکنیم که قسمت وسط میتواند منفی باشد.

مثال ۷ قضیه پردازش داده:[۶] فرض کنید X از کانال $p(y|x)$ رد شده و Y را تولید نماید و Y از کانال $p(z|y)$ رد شده و Z را ایجاد کند. در این حالت توزیع مشترک متغیرهای X و Y و Z بشکل زیر است:

$$p(x,y,z)=p(x)p(y|x)p(z|y).$$

قضیه پردازش داده بیان میدارد که

$$I(X;Y)\geq I(X;Z).$$

اثبات: با استفاده از توزیع مشترک متغیرهای X و Y و Z میتوان نشان داد که

$$p_{X,Z|Y}(x,z|y)=p_{X|Y}(x|y)p_{Z|Y}(z|y)$$

و در نتیجه

$$I(X;Z|Y)=\sum_{x,y,z}p_{X,Y,Z}(x,y,z)\log\frac{p_{X,Z|Y}(x,z|y)}{p_{X|Y}(x|y)p_{Z|Y}(z|y)}=0.$$

یا اینکه X و Z بشرط دانستن Y از هم مستقل هستند. پس یکی از نواحی نمودار ون صفر است. عبارت

$$I(X;Y)-I(X;Z)$$

برابر است با تفاضل دو اطلاعات مشترک که با علامت مثبت و منفی آنها را در شکل ۷ مشخص کرده ایم. از روی شکل ۷ واضح است که قسمت وسط ساده شده و قسمت منفی سمت چپ صفر است، پس نامساوی برقرار است.

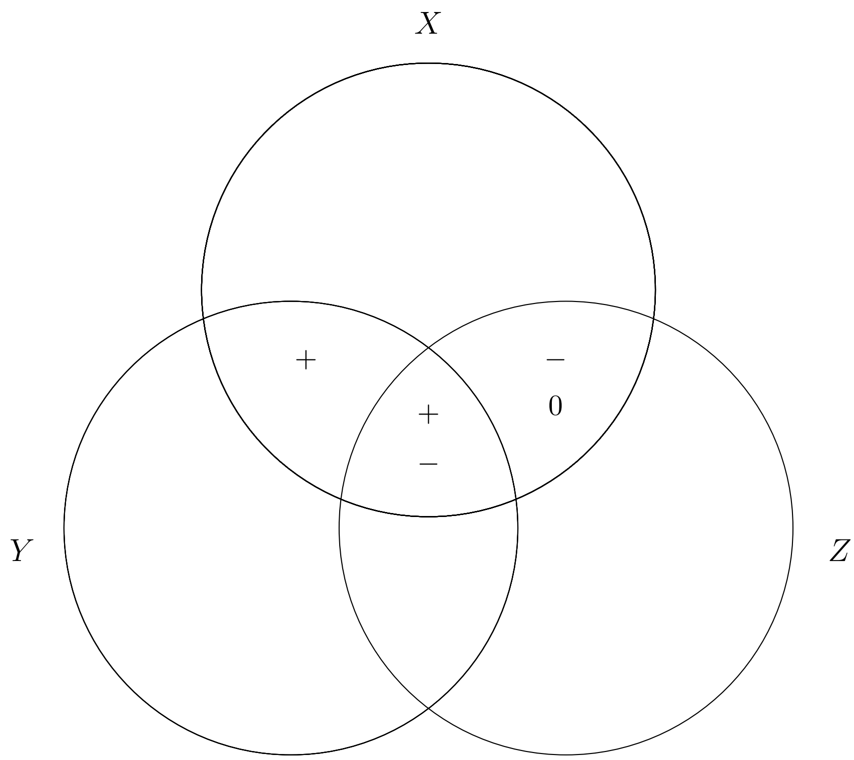
نکته ۸ اگر توزیع مشترک متغیر های X و Y و Z بشکل

$$p(x,y,z)=p(x)p(y|x)p(z|y)$$

باشد میگوییم که متغیرهای X و Y و Z تشکیل یک زنجیره مارکف میدهند و مینویسیم $X\to Y\to Z$.

تمرین ۹ با استفاده از نمودار ون تحقیق کنید که

[۶] Data Processing Theorem



شکل ۵: نمودار ون برای سه متغیر در مثال مربوط به قضیه پردازش داده.

- $I(X;Y|Z) \leq H(X)$
- $I(X;Y) + H(Z) \geq I(X;Y|Z) \geq I(X;Y) - H(Z)$
- $H(X|Y) + H(Y|Z) \geq H(X|Z)$
- اگر Z تابعی از (X,Y) باشد و همچنین X از Z مستقل باشد آنگاه $H(Y) \geq H(Z)$

۲.۵ نامساوی های آنتروپی با چهار متغیر و بیشتر

در حالت کلی تر که 4 متغیر داشته باشیم می توان عبارات را بر حسب آنتروپی متغیرها بسط داد. یعنی هرچه عبارت اطلاعات مشترک داریم بر حسب آنتروپی بست دهیم و آن را باز کنیم. در این صورت به عبارتی میرسیم که تماما شامل جملات آنتروپی است. مثلا

$$2H(X) - 4H(X,Z) + H(X,Y,Z,T) - 8H(X,Y,T) + 9H(T) - H(Y) - H(Y,T) \geq 0$$

برای اینکه بدانیم روابطی مانند رابطه بالا برقرار است یا نه، از سری روابط زیر که به نامساویهای از نوع شانون مشهور است استفاده می شود:

$$H(X) \geq 0$$

$$H(X,Y) \geq H(X)$$

$$H(X,Y) \leq H(X) + H(Y)$$

$$I(X;Y|Z) \geq 0$$

نامساوی های از نوع شانون را برای تمامی متغیرهایی که داریم می نویسیم و لیستی بلند از تمامی نامساوی های از این دست حاصل می شود. سپس سعی میکنیم که با ضرب هر کدام از این نامساوی ها در یک عدد مثبت و جمع کردن آنها نامساوی مورد دلخواه خود را ثابت کنیم. نرم افزاری به نام $ITIP$ [۷] برای انجام این کار وجود دارد. برخی اوقات ممکن است یک رابطه را نتوان بر اساس نامساوی های فوق ثابت نمود لذا نمی توان در مورد مثبت و یا منفی بودن آن صحبت نمود. نامساوی های از نوع غیرشانون نیز وجود دارند که از حوصله این بحث خارج است.

۳.۵ نامساوی فانو و پیوستگی آنتروپی

اگر $Y = f(X)$ باشد، $H(Y|X) = 0$ است. از روی پیوستگی حدس میزنیم که اگر Y تقریبا تابعی از X باشد، آنوقت $H(Y|X) \approx 0$. اینکه تقریبا تابعی از X باشد را میتوان اینگونه بیان کرد که $P\big(Y = f(X)\big) \geq 1 - \epsilon$ برای تابعی مانند f برقرار است.

یک متغیر تصادفی T برنولی به صورت تابع مشخصه در نظر می گیریم.

$$T = \mathbf{1}[f(X) = Y]$$

[۷] http://user-www.ie.cuhk.edu.hk/ ITIP/

دقت کنید که

$$P(T=1) \geq 1-\epsilon.$$

حال داریم

$$\begin{aligned} H(Y|X) &\leq H(T,Y|X) \\ &= H(T|X) + H(Y|X,T) \\ &\leq H(T) + H(Y|X,T) \\ &= h(\epsilon) + P(T=0)H(Y|X,T=0) + P(T=1)H(Y|X,T=1) \\ &= h(\epsilon) + P(T=0)H(Y|X,T=0) \\ &\leq h(\epsilon) + \epsilon \log\left(|\mathcal{Y}|\right) \end{aligned}$$

تمرین ۱۰ ثابت کنید شکل قویتر زیر از نامساوی فانو نیز برقرار است

$$H(Y|X) \leq h(\epsilon) + \epsilon \log\left(|\mathcal{Y}|-1\right)$$

نکته ۱۱ سمت راست نامساوی فانو هیچ ربطی به الفبای X ندارد و فقط به الفبای Y ربط دارد.

تمرین ۱۲ ثابت کنید باندی قوی تر از آنچه در تمرین ۱۰ آمده نخواهیم داشت. یعنی میتوان متغیرهای X و Y را بگونه ای یافت که $P\big(Y=f(X)\big) \geq 1-\epsilon$ برای تابعی مانند f و در عین حال

$$H(Y|X) = h(\epsilon) + \epsilon \log\left(|\mathcal{Y}|-1\right)$$