نظریه اطلاعات کوانتمی ۱، ترم پاییز ۱۳۹۲-۱۳۹۱

مدرسین: ابوالفتح بیگی و امین زاده گوهری

مروری بر تئوری اطلاعات کلاسیک

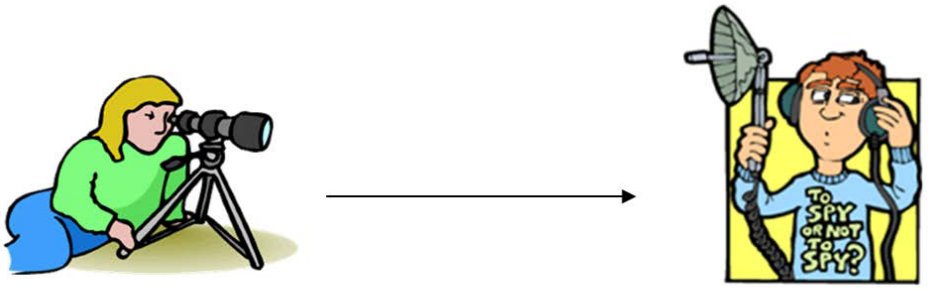
هدف از این متن مروری بر تئوری اطلاعات کلاسیک است. با این که تلاش شده که برخی از مفاهیم پایه‌ای خلاصه شوند، این متن کافی نبوده و ضروری است به کتب مربوطه که به شکل مبسوط‌تری به موضوع پرداخته‌اند نیز مراجعه شود.

۱ اهمیت کدگذاری بلوکی

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:

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

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

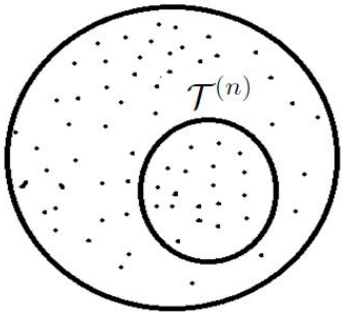
برای بیان مفهوم نوعی بودن با آزمایش پرتاب سکه شروع می‌کنیم که می‌توان آن را با یک متغیر تصادفی برنولی مدل کرد. فرض کنید متغیر برنولی‌ای داریم که توزیع احتمال آن به شکل زیر است.

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

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:

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

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:



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.

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

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 $n > N$

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

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

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

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

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

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 x^n مانند 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 $n\pi(1|x^n)$.) 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

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

روابط (۱) و (۲) نتیجه می‌دهند که

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

پس طبق باند مجموع[1]

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

یا

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

یعنی اگر مجموعه دنباله‌هایی را در نظر بگیریم که

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

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 a 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.

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 $\binom{n}{\frac{n}{2}}\approx 2^n$ است. در واقع در بسط

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

[1] Union Bound

دیگر جملات در مقابل $\binom{n}{\frac{n}{2}}$ قابل صرف نظر کردن هستند. رابطه $\binom{n}{\frac{n}{2}} \approx 2^n$ 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: $\binom{n}{\frac{n}{2}} \times (\frac{1}{2})^n$. 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:

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

۲.۱.۲ حالت کلی

در حالت کلی وقتی X دودویی نیست، مجموعه نوعی این گونه تعریف می‌شود:

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

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

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

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

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

که در آن

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

به این عبارت آنتروپی (شانون) گفته می‌شود.

نکته ۳ دو روش متفاوت برای تعریف دنباله‌های نوعی در متون تئوری اطلاعات وجود دارد که به نوعی بودن قوی و ضعیف[۲] معروف هستند. ما در اینجا تنها در مورد مفهوم نوعی بودن قوی صحبت می‌کنیم. دلیل اصلی آن سادگی بحث و مفهوم شهودی دنباله‌های نوعی قوی می‌باشد. 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.

۲.۲ دو یا چند متغیر

مجموعه‌های نوعی را برای بیش از دو متغیر هم می‌توان تعریف کرد:

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

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

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

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

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 $(x=0,y=1)$ ها np_{01} و تعداد $(x=1,y=0)$ ها np_{10} 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 باشد هرکدام از x^n و y^n نیز (به تنهایی) نوعی است زیرا به عنوان مثال تعداد صفرها در دنباله x^n برابر است با:

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

۳.۲ نوعی بودن شرطی

نوعی بودن شرطی 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:

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

[۲] Strong and weak typicality

حال اگر دنباله‌ی نوعی $x^n = 0100110\cdots0$ 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

$$\begin{aligned} x^n =& 0\ 1\ 0\ 0\ 1\ 1\ 0\cdots\ 0 \\ & 0\quad 0\,0\quad\ 0\cdots 0 \\ & 1\quad\ \ 1\,1\ \ \cdots \end{aligned}$$

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:

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

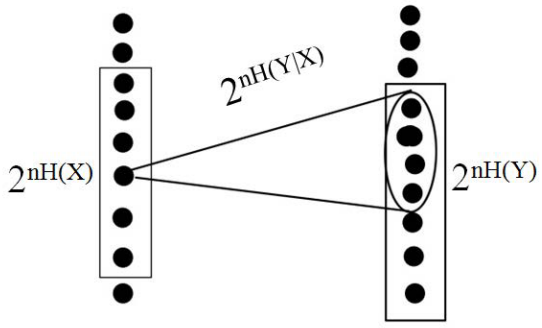
دقت کنید که رابطه بالا توجیهی برای تعریف $H(Y|X) := H(X,Y) - H(X)$ است.

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

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



شکل ۲: گراف نوعی

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 ۵ 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

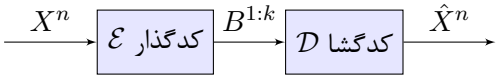
$$P\big((X^n,Y^n)\in\mathcal{T}_\epsilon^{(n)}\big) = \sum_{(x^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} p(x^n,y^n).$$

۴.۲ گراف نوعی

گراف نوعی[۳] 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 $2^{nH(Y|X)}$

[۳] Typicality Graph



شکل ۳: نمایش شماتیک یک کدگذار منبع

دنباله با x^n نوعی هستند پس

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

۳ کدگذاری منبع

زمانی که می‌خواهیم مشاهدات خود از یک منبع .i.i.d با توزیع $p(x)$ را فشرده کنیم از کدگذاری منبع استفاده می‌کنیم. هر کدی با یک سری از پارامترها مشخص می‌شود. اولین پارامتر یک کد، طول کد، n، است. پارامتر دومی که باید تعیین شود، نرخ فشرده سازی کد، R، است. یعنی پس از مشاهده n نسخه از منبع $x^n \in \mathcal{X}^n$، می‌خواهیم آن را در $k = nR$ بیت فشرده کنیم و برای یک گیرنده ارسال کنیم. شکل ۳ نمایش شماتیک یک کدگذار منبع را نشان می‌دهد.

برای فشرده‌سازی $x^n \in \mathcal{X}^n$ نیاز به یک کدگذار داریم. وظیفه کدگذار تبدیل x^n به دنباله‌ای $k = nR$ بیتی است.

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

همچنین نیاز به یک تابع کدگشا داریم تا سمبل‌های اصلی را پس از فشرده‌سازی بازیابی کنیم.

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

چهارتایی $(n, R, \mathcal{E}, \mathcal{D})$ یک کد منبع را مشخص می‌کند.

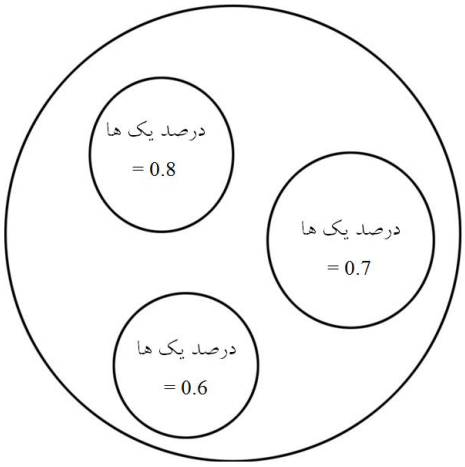
مفهوم بعدی که باید تعریف شود، احتمال خطای یک کد است. دقت کنید که احتمال خطای یک کد جزو پارامترهای تعریف آن نیست، بلکه چیزی است که پس از تعریف یک کد قابل محاسبه است.

$$P(\text{خطا}) = P(\mathcal{D} \circ \mathcal{E}(X^n) \neq X^n) = \sum_{x^n : \mathcal{D} \circ \mathcal{E}(x^n) \neq x^n} p(x^n).$$

فرض کنید قرار است کدی طراحی شود که احتمال خطای آن از $\epsilon = 10^{-4}$ کمتر باشد. از طرفی این حق برای طراح کد مفروض است که طول کد n را تا آنجا که می‌خواهد زیاد کند. این بدان معنی است که با یک مقدار خطای ϵ داده شده (در اینجا همان مقدار 10^{-4}) و $n \to \infty$، باید به کدهایی توجه شود که احتمال خطای آنها کمتر از ϵ و نرخ فشرده‌سازی آنها از همه کمتر باشد. به این نرخ کمینه C_ϵ می‌گوییم:

$$C_\epsilon := \inf_{P(\text{خطا}) < \epsilon} R.$$

در رابطه فوق C_ϵ کمترین فشرده‌سازی است به گونه‌ای که خطای بازیابی حداکثر ϵ باشد.



شکل ۴: افراز دنباله های دودویی به طول n بر حسب درصد یک‌های دنباله

در این صورت می‌توان نرخ فشرده‌سازی را این‌گونه تعریف کرد:

$$C := \lim_{\epsilon \to 0} C_\epsilon.$$

در صورتی که یک منبع i.i.d. با توزیع $p(x)$ داشته باشیم بهترین نرخ فشرده‌سازی آن $C = H(X)$ است. جهت اثبات قابل حصول بودن کافی است که توجه کنیم که تعداد دنباله‌های نوعی حدودا $2^{nH(X)}$ است. . فرض کنید که این دنباله‌ها را به ترتیب با اعداد $1, 2, \cdots, 2^{nH(X)}$ شماره‌گذاری کرده باشیم. در این صورت کافی است که کدگذار در صورتی که دنباله منبع نوعی بود شماره آن به عنوان فشرده‌سازی آن در نظر بگیرد. و در صورتی که دنباله مشاهده شده از منبع نوعی نباشد، به صورت تصادفی یک عدد بین 1 تا 2^{nR} به عنوان مقدار فشرده شده انتخاب کند. از آن جایی که با احتمال زیاد دنباله مشاهده شده نوعی است، احتمال خطای این کدگذار کوچک و در حد $n \to \infty$ به سمت صفر می‌رود. پس نرخ فشرده‌سازی $H(X)$ قابل حصول است.

جهت اثبات ضروری بودن این نرخ فشرده‌سازی کافی است از نامساوی فانو و پردازش داده استفاده کنیم:

$$nH(X) \approx I(X^n; \hat{X}^n) \leq H(B^{1:nR}) \leq nR$$

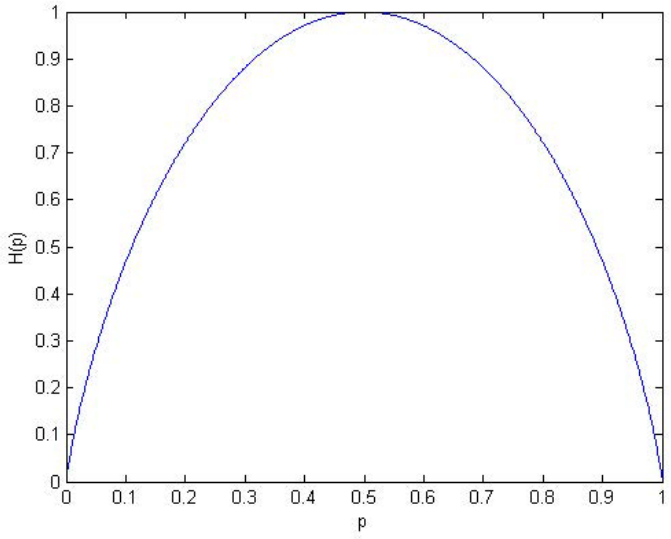
که در آن $B^{1:nR}$ همان دنباله فشرده شده از بیت‌ها است.

۴ آنتروپی نسبی

۱.۴ افراز تمامی دنباله‌ها و محاسبه احتمال دسته‌هایی از دنباله‌های غیرنوعی

می توان مجموعه تمامی دنباله‌های دودویی n-تایی را به مجموعه‌هایی افراز نمود که $0.9n$ ، $0.8n$ ، $0.7n$ و $\cdots$ یک دارند. شکل ۴ را ببینید.

دقت کنید که هنوز صحبت از هیچ متغیر تصادفی‌ای نکرده‌ایم، بلکه فقط مجموعه دنباله‌ها را افراز کرده‌ایم. تعداد



شکل ۵: تابع آنتروپی باینری.

اعضای مجموعه‌های با تعداد یک های $0.9n$ و $0.8n$ برابر هستند با:

$$\binom{n}{0.9n} \approx 2^{nh(0.9)},$$

$$\binom{n}{0.8n} \approx 2^{nh(0.8)},$$

که در اینجا از تقریب استرلینگ استفاده شده است و تابع

$$p \mapsto h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$$

همان تابع آنتروپی دودویی است. نمودار $h(p)$ در شکل ۵ آمده است.

حال اگر سکه‌ای در نظر گرفته که با احتمال 0.9 حاصل آن پشت است، و آن را n بار پرتاب کنیم، دنباله‌ی حاصل با احتمال نزدیک به 1 در مجموعه "درصد یک‌ها 0.9" قرار می گیرد. اما دقت کنید که احتمال وقوع هر دنباله‌ای هر قدر هم که کوچک، ممکن است ناصفر باشد. در واقع با احتمال خیلی کمی ممکن است که دنباله حاصل در خارج از این مجموعه و مثلا در مجموعه "درصد یک‌ها 0.8" قرار بگیرد. به عبارت دیگر با احتمال خیلی کمی ممکن است سکه ما مانند سکه‌ای عمل کند که احتمال یک آمدن آن 0.8 باشد.

سکه‌ای که احتمال پشت آمدن آن برابر 0.9 است به صورت نسبی محتمل‌تر است که شبیه سکه‌ای با احتمال پشت آمدن 0.8 عمل کند تا سکه‌ای که احتمال پشت آمدن آن 0.7 باشد. هر چه درصد یک‌ها به هم نزدیک‌تر باشد، عبور از یکی به دیگری محتمل‌تر خواهد بود. لذا فاصله میان درصد یک‌ها مهم خواهد بود.

حال احتمال این که سکه‌ی با پارامتر 0.9 همانند سکه‌ای با پارامتر 0.8 رفتار کند را حساب می‌کنیم. یعنی احتمال

این که تعداد یک‌ها در پرتاب سکه‌ی با پارامتر 0.9 برابر $0.8n$ باشد. این احتمال برابر است با

$$\begin{aligned}\Pr &= \binom{n}{0.8n}(0.9)^{0.8n}(0.1)^{0.2n}\\ &\approx 2^{nh(0.8)}2^{0.8n\log(0.9)+0.2n\log(0.1)}\\ &= 2^{-n\left(0.8\log\frac{0.8}{0.9}+0.2\log\frac{0.2}{0.1}\right)}\\ &:= 2^{-nD\left(\ (0.8,0.2)\middle\|\ (0.9,0.1)\ \right)}\end{aligned}$$

که در آن اگر توزیع p را توزیع $(0.8, 0.2)$ و توزیع q را توزیع $(0.9, 0.1)$ بگیریم به عبارت

$$D(p\|q) = \sum_x p(x)\log\frac{p(x)}{q(x)}$$

می‌رسیم. به $D(p\|q)$ فاصله کولبک-لیبلر[۴] یا آنتروپی نسبی[۵] دو توزیع p و q گفته می‌شود. این عبارت در واقع «فاصله‌ای» میان دو توزیع تعریف می‌کند. اگر $p = q$ باشد آنگاه $D(p\|q) = 0$ و اگر $p \neq q$ باشد همواره داریم $D(p\|q) > 0$. با این حال این معیار فاصله متقارن نیست. یعنی در حالت کلی $D(p\|q) \neq D(q\|p)$.

با توجه به محاسبات بالا می‌بینیم که هر چه $D(p\|q)$ بیشتر باشد احتمال اینکه نمونه‌های توزیع q خودشان را به عنوان نمونه‌های توزیع p جا بزنند کمتر خواهد بود. دقت کنید که با افزایش n احتمال $2^{-nD(p\|q)}$ به هر حال به سمت صفر میل می‌نماید چون احتمال مجموعه نوعی به سمت 1 باید برود (و بنابراین احتمال مجموعه غیر نوعی به سمت 0). اما هر چه $D(p\|q)$ بیشتر باشد «سرعت به سمت صفر رفتن» این احتمال بیشتر خواهد بود.

نکته ۶ یک توضیح در مورد نمادگذاری: در عبارت بالا صحبت از $p(x)$ و $q(x)$ کردیم. فرض کنید که الفبای 0 و 1 را داریم، و می‌خواهیم در آن واحد راجع به دو توزیع مختلف روی الفبای 0 و 1 صحبت کنیم، مثلا یک توزیع که به 1 احتمال 0.9 می‌دهد و توزیع دیگر که به 1 احتمال 0.8 می‌دهد. جهت انجام این کار دو راه وجود دارد. یکی اینکه دو متغیر تصادفی تعریف کنیم، مثلا X_1 و X_2 که توزیع X_1 برابر با $(0.8, 0.2)$ و توزیع X_2 برابر با $(0.9, 0.1)$ باشد. اما راه دیگر این است که از دو تابع وزن[۶] مختلف روی الفبای $\mathcal{X}$ استفاده کنیم.

$$p_X(0) = 0.1, p_X(1) = 0.9,$$

$$q_X(0) = 0.2, q_X(1) = 0.8.$$

دقت کنید که متغیر تصادفی X به خودی خود معنی ندارد مگر اینکه بگوییم که آن را با تابع وزن p استفاده می‌کنیم یا با تابع وزن q. در بحث قبلی که صحبت از $p(x)$ و $q(x)$ کردیم منظور همین دو تابع وزن مختلف بود که روی مجموعه یکسان $\mathcal{X}$ تعریف شده‌اند.

[۴] Kullback–Leibler divergence
[۵] Relative Entropy
[۶] Measure

۵ شانس تولید وابستگی از نمونه‌های مستقل

فرض کنید آذر و بابک بخواهند که از توزیع احتمال مشترک X, Y زیر نمونه تولید کنند: (در اینجا هدف آذر انتقال پیام به بابک نیست، بلکه این دو صرفا می‌خواهند از یک توزیع احتمال مشترک نمونه تولید کنند)

$$\begin{pmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{pmatrix}$$

می‌بینیم که متغیرهای X و Y داده شده در بالا از هم مستقل نیستند. پس اگر قرار است که آذر نمونه‌های X و بابک نمونه‌های Y را شبیه سازی کنند نیاز به انتقال اطلاعات از طریق یک کانال میان آن دو است. حال فرض کنید که آنها بخواهند بدون استفاده از کانال ارتباطی از توزیع احتمال مشترک نمونه تولید کنند ولی با احتمال خطا. مشخص است که آنها این کار را نمی‌توانند با احتمال بالا انجام دهند، اما با در نظر گرفتن خطای زیاد امکان این کار وجود دارد. (در احتمالات همه چیز ممکن است! مثلا ممکن است تمام مولکولهای هوای موجود در اتاق در یک گوشه اتاق جمع شوند. با احتمال خیلی خیلی کمی این اتفاق می افتد.)

می‌خواهیم احتمال اینکه آنها از توزیع احتمال مشترک نمونه تولید کنند را بدست آوریم. از روی جدول بالا توزیع‌های احتمال حاشیه‌ای X و Y برابراند با $0.3, 0.7$ و $0.4, 0.6$. برای آذر و بابک بهینه است که از همین توزیع‌ها برای تولید نمونه‌های تصادفی استفاده کنند. چون کانال ارتباطی بین این دو وجود ندارد پرتاب سکه‌ها توزیع مستقل برای هر دو خواهد داشت و بهترین انتخاب این توزیع‌های مستقل همان توزیع‌های حاشیه‌ای هستند. توزیع مشترک تولید شده توسط آنها با این فرض برابر با $p_X(x)p_Y(y)$ و جدول توزیع زیر برای پرتاب آنها حاصل می شود.

$$\begin{pmatrix} 0.12 & 0.18 \\ 0.16 & 0.42 \end{pmatrix}$$

احتمال اینکه آذر و بابک از توزیع درست نمونه برداری کنند معادل است با احتمال اینکه توزیع $p(x)p(y)$ مانند توزیع $p(x, y)$ عمل کند (خودش را مانند دیگری جا بزند). از بحث قبل می‌دانیم این احتمال برابر است با:

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

که در آن $I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)$.

با روش زیر هم می‌توانستیم به نتیجه‌ی فوق برسیم: اگر دنباله X^n به صورت i.i.d. از $p_X(x)$ تولید شود و دنباله Y^n نیز i.i.d. از $p_Y(y)$ تولید شود احتمال اینکه X^n و Y^n مشترکا نوعی شوند برابر است با $2^{-nI(X;Y)}$.

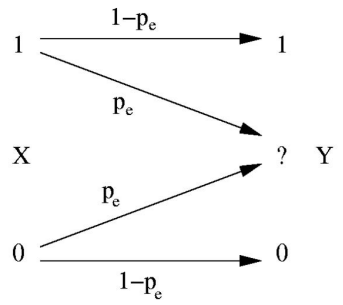
می‌بینیم که هرچه میزان $I(X;Y)$ بیشتر باشد متغیرهای X و Y وابستگی بیشتری داشته و احتمال این که بدون هیچ کانال مخابراتی به صورت تصادفی قابل تولید باشند کمتر می‌شود. اما اگر X و Y مستقل باشد بدون هیچ کانال مخابراتی می‌توان نمونه‌های مشترک آنها را تولید کرد.

۶ کدگذاری کانال

۱.۶ مفهوم کدگذاری تصادفی

پیش از وارد شدن به مبحث کدگذاری کانال با چند مثال شروع می‌کنیم. هدف انگیزه دادن به مفهوم کدگذاری تصادفی است (که ممکن است در نگاه اول عجیب به نظر برسد).

مثال ۷ کانال زیر را با $p_e = 0.1$ در نظر بگیرید:



فرض کنید که دنباله زیر بطول 10 را از طریق این کانال ارسال کنیم.

$$0\,1\,0\,1\,0\,\cdots\,0$$

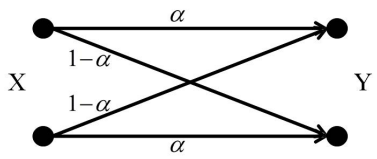
و در نتیجه دنباله دریافتی می تواند بصورت زیر باشد:

$$0\,1\,E\,1\,0\,\cdots\,0$$

فرض کنید که پیش از ارسال به شما گفته شده باشد که از 10 بیت، دقیقا 9 بیت آن سالم به مقصد می‌رسد و یک بیت پاک خواهد شد. سوال این است که چگونه اطلاعات را انتقال دهیم زمانی که نمی‌دانیم که کدام یک از بیت‌ها پاک خواهد شد؟

یک راه حل ساده ارسال 9 بیت اطلاعات و سپس ارسال جمع XOR این 9 بیت است. اما راه دیگر ارسال معادلات XOR تصادفی است، به این معنی که به صورت تصادفی تعدادی از بیت‌ها را گرفته و XOR می‌کنیم و بر روی کانال ارسال می‌کنیم. اگر گیرنده 9 بیت را دریافت کند، 9 معادله خطی (در میدان $\mathbb{F}_2$) را دریافت کرده است. از روی آنها می‌تواند دستگاه 9 معادله 9 مجهول را تشکیل داده و آن را حل کند. اینکه کدام 9 معادله را دریافت کرده مهم نیست تا زمانی که دستگاه معادلات حاصل جواب یکتا داشته باشد. می‌توان ثابت کرد که اگر معادلات به صورت تصادفی تولید شده باشند، با احتمال زیاد معادلات دریافتی از هم مستقل خطی خواهند بود و از روی آنها می‌توان جواب معادلات را بصورت یکتا یافت. این موضوع توجیحی برای استفاده از کدگذاری تصادفی فراهم می آورد.

مثال ۸ کانال BSC زیر را با پارامتر $\alpha = 0.9$ در نظر بگیرید:



جهت ارسال روی این کانال باید از یک کد استفاده کنیم. مثلا بجای ارسال 0 چندین 0 و بجای ارسال 1 چندین 1 ارسال یا از دیگر کدهای معروف استفاده می‌کنیم. اگر در این کانال بخواهیم کلمات کد را انتخاب کنیم، هرچه فاصله همینگ

کلمات کد بیشتر باشد، کد بهتری طراحی شده است. چرا که هنگام عبور از کانال بیتهای اطلاعات دچار خطا می‌شوند، بعضی صفرها به یک تبدیل می‌شوند و بالعکس. به همین دلیل هرچه فاصله کلمات کد از هم بیشتر باشد، قابلیت کشف و تصحیح خطا بیشتر خواهد بود. برای یک کد معیار فاصله‌ی همینگ کمینه بین کلمات کد را در نظر می گیریم. کدی مناسب است که دارای $d_{\min}$ بزرگتری باشد. اما گاهی برای سادگی تحلیل بجای فاصله‌ی کمینه، میانگین فاصله دوبدوی کلمه کدها در نظر گرفته می‌شود.[۷] سوالی که مطرح می‌شود این است که تا چه حد می‌توان کلمات کد را از هم دور کرد به گونه ای که میانگین فاصله دو به دوی این کلمات کد حداکثر شود.

اگر دو کلمه کد بخواهیم انتخاب کنیم، بالطبع کلمه کدهای $000\cdots00$ و $111\cdots11$ را انتخاب می‌کنیم. اگر طول کد n باشد نهایت فاصله‌ی دو کلمه کد n است. اگر سه کلمه کد انتخاب کنیم میانگین فاصله دوبدوی کلمه کدها حداکثر $\frac{2n}{3}$ می شود چرا که از یک طرف در مولفه i-ام، حداقل دو کلمه کد از سه کلمه کد دارای مقدار مساوی هستند. در نتیجه میزان مشارکت مولفه i-ام در متوسط فاصله همینگ دو به دوی کلمات حداکثر $\frac{2}{3}$ است. از طرف دیگر می‌توانیم سه کلمه کد زیر را در نظر بگیریم و به مقدار متوسط فاصله $\frac{2n}{3}$ برسیم.

$$\underbrace{000\cdots00}_{\frac{n}{3}}\underbrace{111\cdots11}_{\frac{n}{3}}\underbrace{111\cdots11}_{\frac{n}{3}}$$
$$\underbrace{111\cdots11}_{\frac{n}{3}}\underbrace{000\cdots00}_{\frac{n}{3}}\underbrace{111\cdots11}_{\frac{n}{3}}$$
$$\underbrace{111\cdots11}_{\frac{n}{3}}\underbrace{111\cdots11}_{\frac{n}{3}}\underbrace{000\cdots00}_{\frac{n}{3}}$$

در نهایت اگر m کلمه کد داشته باشیم و m بزرگ باشد ماکزیمم میانگین فاصله دو به دوی کلمات کد برابر با $\frac{n}{2}$ می‌شود. نشان می‌دهیم که از با کدگذاری تصادفی می‌توان به این کران بالای $\frac{n}{2}$ رسید. اگر m کلمه کد $w_1, w_2, \cdots, w_m$ داشته باشیم، میانگین متوسط فاصله دو به دوی آنها برابر است با

$$\frac{1}{\binom{m}{2}}\sum_{i\neq j} d(w_i, w_j)$$

که در آن $d(\cdot,\cdot)$ فاصله همینگ میان دو کلمه کد است. جمع فاصله‌های همینگ دو به دوی کلمات کد برابر است با جمع تعداد جاهایی که با هم متفاوت هستند. درایه اول تمامی کلمات کد را در نظر بگیرید. جمع فاصله دو به دوی آنها را حساب می‌کنیم. فرض کنید که تعداد $m\beta$ بیت 0 و $m(1-\beta)$ بیت 1 داشته باشیم. اگر دو درایه را انتخاب کرده تنها زمانی مشارکتی در فاصله همینگ خواهیم داشت که یکی از 0-ها و یکی از 1-ها را انتخاب کرده باشد. پس مشارکت این بیت خاص در میانگین برابر است با:

$$\frac{m\beta\cdot m(1-\beta)}{\binom{m}{2}}$$

رابطه فوق زمانی که $\beta=\frac{1}{2}$ باشد ماکزیمم خواهد شد و این بدان معنی است که نصف درایه‌ها در مولفه اول 0 و نصف آنها 1 باشد. در نتیجه کران بالایی میانگین فاصله دو به دوی کدها به صورت زیر می‌شود:

$$\frac{m\beta\cdot m(1-\beta)}{\binom{m}{2}}\cdot n \leq \frac{n}{2}.$$

[۷] البته در طراحی کد بهینه باید حداقل فاصله بین کلمات کد را مورد توجه قرار داد، اما ما در اینجا برای سادگی متوسط فاصله دو به دو را در نظر می‌گیریم.

حال نکته جالب این است که از طریق کدگذاری تصادفی می‌توان به این کران بالای $\frac{n}{2}$ رسید. در واقع انتخاب تصادفی کلمات کد باعث می‌شود که کلمات کد بخوبی پخش شده و فاصله آنها از هم زیاد شود. فرض کنید که برای تعیین کلمات کد، از پرتاب سکه متقارن استفاده کنیم. یعنی برای مشخص کردن هر مولفه هر کلمه کد این گونه عمل کنیم: هربار که نتیجه "شیر" آمد از 0 و هربار که نتیجه "خط" آمد از 1 در بیت‌ها استفاده کنیم. در این صورت بنابر قانون اعداد بزرگ در هر بیت خاص نیمی صفر و نیمی یک خواهیم داشت و لذا به کران بالای $\frac{n}{2}$ می‌رسیم.

۲.۶ کدگذاری کانال

هدف کدگذاری کانال، انتقال اطلاعات بر روی یک کانال مخابراتی است.

کانال مخابراتی: فرض کنید یک کانال دلخواه با الفبای ورودی $\mathcal{X}$ و الفبای خروجی $\mathcal{Y}$ داریم.[۸] برای مشخص کردن یک کانال نیاز داریم که ابتدا الفبای ورودی، سپس الفبای خروجی و در نهایت ضابطه کانال $p(y|x)$ (ویا اصطلاحاً توزیع خروجی به شرط ورودی) را مشخص کنیم. یعنی برای مشخص کردن یک کانال از سه‌تایی $\big(\mathcal{X}, \mathcal{Y}, p(y|x)\big)$ استفاده می‌کنیم. به عنوان مثال، برای یک کانال BEC، $\mathcal{X} = \{0, 1\}$ و $\mathcal{Y} = \{0, E, 1\}$.

تعداد دفعات زیاد استفاده از کانال مخابراتی: مشابه کدگذاری منبع، اگر به جای اینکه یک بار از کانال استفاده کنیم، به تعداد دفعات زیاد از کانال استفاده کنیم، می‌توانیم به نرخ‌های بهتر (حداقل با خطای کمتر) برای انتقال اطلاعات برسیم. اما زمانی که داده به صورت بلوکی در کانال منتقل می‌شود، هزینه این انتقال تاخیر برای سیستم خواهد بود. اما در تئوری اطلاعات، و در تعاریف ظرفیت، توجهی به این تاخیرها نمی‌شود.

تعریف کد: زمانی که می‌خواهیم اطلاعات را بر روی کانال ارسال کنیم، باید یک کد بسازیم. هر کدی با یک سری پارامتر مشخص می‌شود. اولین پارامتر یک کد، طول کد، n، است. پارامتر دومی که باید تعیین شود، نرخ کد، R، است که نشان می‌دهد به ازای n بار استفاده از کانال می‌خواهیم $k = nR$ بیت منتقل کنیم. یک پیام nR بیتی را می‌توان با یک دنبالی از $\{0,1\}^{nR}$ و یا یک عدد در بازه $\{1, 2, 3, \cdots, 2^{nR}\}$ نمایش داد. این دو نمایش معادل هستند. ما نمایش دوم را برمی‌گزینیم.

برای انتقال پیام $m \in \{1, 2, 3, \cdots, 2^{nR}\}$ نیاز به یک کدگذار داریم. وظیفه کدگذار تبدیل پیام m به دنباله‌ای از ورودی‌های کانال است. یعنی به ازای هرپیام، یک کلمه کد n بیتی تولید می‌شود. در نتیجه جدول کلمات کد را می‌توان به صورت زیر تشکیل داد. سطرهای این جدول کلمات کد هستند. در این جدول فرض شده که ورودی کانال دودویی است، $\mathcal{X} = \{0, 1\}$. به همین دلیل خانه‌های جدول با 0 و 1 پر شده‌اند.

n	$\cdots$	3	2	1	کلمه کد
1	$\cdots\ \cdots$	0	1	1	1
0	$\cdots\ \cdots$	0	0	1	2
1	$\cdots\ \cdots$	0	1	0	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0	$\cdots\ \cdots$	0	1	0	2^{nR}

[۸] مرسوم است که برای الفبای متغیرهای تصادفی از حروف کالیگرافیک استفاده کنیم.

شکل ۶: نمایش شماتیک یک کدگذار کانال

برای مشخص کردن کد، تابع کدگذار

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

و تابع کدبردار

$$\mathcal{D} : \mathcal{Y}^n \to \{1, 2, 3, \cdots, 2^{nR}\}$$

نیز علاوه بر n و R باید تعیین شوند. پس یک کد با چهارتایی $(n, R, \mathcal{E}, \mathcal{D})$ مشخص می‌گردد.

احتمال خطای کد: مفهوم بعدی که باید تعریف شود، احتمال خطای یک کد است. دقت کنید که احتمال خطای یک کد جزو پارامترهای تعریف آن نیست، بلکه چیزی است که پس از تعریف یک کد قابل محاسبه است. تعاریف احتمال خطای متوسط و احتمال خطای بیشینه برای یک کد وجود دارد. قبل از بیان این دو تعریف به احتمال خطای یک کد به شرط اینکه پیام خاصی ارسال شده باشد می‌پردازیم. احتمال خطای کد به شرط ارسال پیام $m = 1$ را می‌توان به صورت زیر تعریف کرد:

$$P(\text{خطا}|m = 1) = \sum_{y^n : \mathcal{D}(y^n) \neq 1} p(y^n|x^n(1))$$

که منظور از $x^n(1)$ همان کلمه کد متناظر با پیام $m = 1$ است: $x^n(1) = \mathcal{E}(1)$. در معادله فوق، $y^n : \mathcal{D}(y^n) \neq 1$ یعنی "تمام دنباله‌های y^n که ممکن است اتفاق بیافتند به طوری که خروجی کدبردار 1 نشود." مطابق با اصل ضرب و با استفاده از بدون حافظه بودن کانال می‌توان نوشت:

$$p(y^n|x^n(1)) = \prod_{i=1}^{n} q(y_i|x_i(1)).$$

احتمال خطای یک کد را به دو صورت می‌توان تعریف کرد: (۱) احتمال خطای متوسط، (۲) احتمال خطای حداکثر. احتمال خطای متوسط به صورت زیر است:

$$Pe_{ave} = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} P(\text{خطا}|m = i).$$

احتمال خطای بیشینه به صورت زیر تعریف میشود:

$$Pe_{max} = \max_{1 \leq i \leq 2^{nR}} P(\text{خطا}|m = i).$$

از آنجایی که m به صورت یکنواخت و تصادفی از $\{1, 2, 3, \cdots, 2^{nR}\}$ انتخاب می‌شود، می‌توان آن را به صورت یک متغیر تصادفی M در نظر گرفت. در نتیجه x^n نیز به متغیر تصادفی X^n تبدیل می‌شود ($X^n = \mathcal{E}(M)$) . به همین

ترتیب y^n نیز یک متغیر تصادفی به صورت Y^n و خروجی کدبردار نیز یک متغیر تصادفی با نام $\hat{M}$ خواهد بود. این متغیرهای تصادفی در قسمت وارون مساله کدگذاری کانال استفاده خواهند شد.

تعریف ظرفیت: در منابع مختلف، ظرفیت یک کانال مخابراتی به چندین صورت تعریف شده است که همه آنها معادل هم هستند. فرض کنید قرار است کدی طراحی شود که احتمال خطای آن از $\epsilon = 10^{-4}$ کمتر باشد. از طرفی این حق برای طراح کد مفروض است که طول کد n را تا آنجا که می‌خواهد زیاد کند. این بدان معنی است که با یک مقدار خطای ϵ داده شده (در اینجا همان مقدار 10^{-4}) و $n \to \infty$، باید به کدهایی توجه شود که احتمال خطای آنها کمتر از ϵ و نرخ ارسال آنها از همه بیشتر باشد. به این نرخ ماکزیمم C_ϵ می‌گوییم:

$$C_\epsilon := \sup_{Pe<\epsilon} R.$$

در رابطه فوق C_ϵ بیشترین نرخی است که می‌توان روی کانال ارسال کرد به گونه‌ای که خطا حداکثر ϵ باشد. در این صورت ظرفیت یک کانال برابر است با:

$$C := \lim_{\epsilon \to 0} C_\epsilon.$$

با توجه به اینکه از معیار احتمال خطای بیشینه و یا متوسط در تعریف بالا استفاده کنیم به دو تعریف از ظرفیت می‌رسیم. ولی بعدا خواهیم دید که این دو تعریف جواب یکسانی می‌دهند. این جواب به شکل زیر است:

$$C = \max_{p(x)} I(X;Y)$$

که در آن $I(X;Y)$ اطلاعات متقابل بین دو متغیر تصادفی X و Y است.

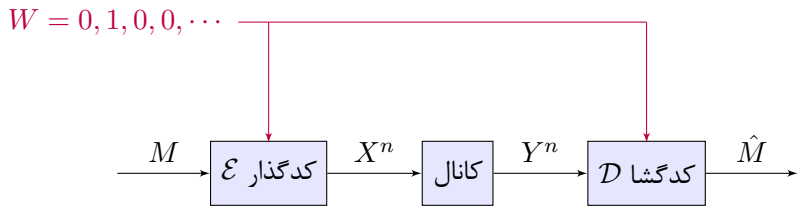
مثال ۹ ظرفیت یک کانال *BSC* را می‌توان به صورت زیر حساب کرد:

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

$$\begin{aligned} H(YX) &= p(X=0)H(YX=0) + p(X=1)H(YX=1) \\ &= h(p)P(X=0) + h(p)P(X=1) \\ &= h(p) \end{aligned}$$

این یعنی اینکه جمله $H(YX)$ به توزیع ورودی وابسته نیست. پس برای حداکثر کردن ظرفیت کانال باید جمله $H(Y)$ حداکثر شود. می‌دانیم حداکثر مقدار $H(Y)$ زمانی به دست می‌آید که توزیع آن یکنواخت باشد و این حداکثر، برابر است با $\log(|\mathcal{Y}|)$. با توجه به اینکه الفبای خروجی دودویی است، پس می‌توان نوشت:

$$C \leq \log(|\mathcal{Y}|) - H(YX) = 1 - h(p).$$



شکل ۷: نمایش شماتیک یک کدگذار کانال به همراه منبع تصادفی به اشتراک گذاشته شده

۳.۶ قسمت مستقیم

منظور از قسمت مستقیم یا قسمت قابل حصول[۹] اثبات این است که نشان دهیم

$$C \geq \max_{p(x)} I(X;Y).$$

ایده اثبات قسمت مستقیم رابطه ظرفیت، ایده کدگذاری تصادفی[۱۰] است. برای سادگی فرض می‌کنیم که کانال دارای ورودی دودویی بوده و توزیع ورودی‌ای که عبارت $I(X;Y)$ را ماکزیمم می‌کند توزیع یکنواخت است. همچنین فرض می‌کنیم گیرنده بتواند کارهای تصادفی انجام دهد. به همین دلیل نیاز به یک منبع تصادفی داریم (منبع نیاز به یک رشته تصادفی از 0 و 1 دارد). فعلاً فرض می‌کنیم یک دنباله تصادفی طولانی W با توزیع یکنواخت از $0/1$ وجود دارد که مستقل از پیام است و بین کدبردار و کدگذار به اشتراک گذاشته می‌شود (مطابق شکل ۷). بعداً نشان داده خواهد شد که این دنباله تصادفی را می‌توان حذف کرد.

کدگذار و کدبردار با استفاده از این رشته تصادفی، کتاب کد را تولید می‌کنند (از آنجایی که W یک رشته تصادفی است، کتاب کد نیز تصادفی خواهد بود). در نتیجه جدول کلمات کد را مطابق جدول زیر با استفاده از مقادیر تصادفی رشته W به ترتیب پر کرده و یک کتاب کد تصادفی می سازند.

n	$\cdots$	3	2	1	کلمه کد
$W(n)$	$\cdots$ $\cdots$	$W(3)$	$W(2)$	$W(1)$	1
$W(2n)$	$\cdots$ $\cdots$	$W(n+3)$	$W(n+2)$	$W(n+1)$	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$W(2^{nR}n)$	$\cdots$ $\cdots$	$W\big((2^{nR}-1)n+3\big)$	$W\big((2^{nR}-1)n+2\big)$	$W\big((2^{nR}-1)n+1\big)$	2^{nR}

کدگذار پیام M را انتخاب کرده و با توجه به کتاب کد، دنباله $X^n(M)$ را ارسال می‌کند. بعد از عبور دنباله از کانال، دنباله Y^n بدست خواهد آمد. انتظار می رود که دنباله Y^n با دنباله $X^n(M)$ نوعی باشد. زیرا زمانی که ورودی کانال نوعی است، انتظار داریم که خروجی کانال نیز با احتمال زیاد با آن نوعی باشد. در این زمان گیرنده کتاب کد را جستجو کرده و تمامی کلمات کد را که با Y^n نوعی هستند را پیدا می کند. چند حالت ممکن است اتفاق بیافتند: (۱) این لیست تهی است، (۲) این لیست شامل تنها یک دنباله است که در این صورت همان دنباله را در خروجی قرار می‌دهد (۳) و یا

[۹] Achievability part or the Direct part
[۱۰] Random Coding

این لیست شامل بیشتر از یک دنباله است. برای محاسبه احتمال خطا، فرض می‌کنیم پیام $M = 1$ ارسال شده است (به دلیل وجود تقارن، تفاوتی بین پیام‌ها برای محاسبه احتمال خطا وجود ندارد). در نتیجه می‌توان نوشت:

$$\begin{aligned} P(\text{خطا}|M = 1) =& P(\text{ هیچ کلمه کدی با } Y^n \text{ نوعی نباشد } |M = 1) + \\ & P(\text{ حداقل دو دنباله با } Y^n \text{ نوعی باشد } |M = 1) + \\ & P(\text{ یک دنباله اشتباه با } Y^n \text{ نوعی باشد } |M = 1) + \end{aligned}$$

احتمال "هیچ کلمه کدی با نوعی نباشد" خیلی کوچک است و به سمت صفر میل می‌کند، زیرا زمانی که یک دنباله نوعی از کانال $p(y|x)$ عبور می‌کند، با احتمال زیاد y^n با دنباله ورودی x^n نوعی است. بنابراین با احتمال زیاد پیام ارسالی در لیست گیرنده خواهد بود. برای محاسبه احتمال بقیه خطاها (خطای اینکه دنباله اشتباهی در لیست گیرنده قرار بگیرد)، یک سری واقعه به صورت زیر تعریف می‌کنیم:

$$\begin{aligned} A_2 &: \text{واقعه ای که کلمه کد 2 با } Y^n \text{ نوعی باشد} \\ A_3 &: \text{واقعه ای که کلمه کد 3 با } Y^n \text{ نوعی باشد} \\ &\cdots \\ A_{2^{nR}} &: \text{واقعه ای که کلمه کد } 2^{nR} \text{ با } Y^n \text{ نوعی باشد} \end{aligned}$$

حال با استفاده از باند مجموع می‌توان نوشت:

$$\begin{aligned} P(Error|M = 1) &\approx P(A_2 \cup A_3 \cup \cdots \cup A_{2^{nR}}) \\ &\leq P(A_2) + P(A_3) + \cdots + P(A_{2^{nR}}) \\ &= (2^{nR} - 1)P(A_2) \end{aligned}$$

قبلا دیده بودیم که اگر دو دنباله را بصورت مستقل از توزیع حاشیه‌ای تولید کنیم احتمال اینکه دو دنباله مشترکا نوعی شوند برابر است با $P(A_2) \approx 2^{-nI(X;Y)}$. در نتیجه:

$$P(\text{خطا}|M = 1) \leq 2^{nR}2^{-nI(X;Y)} \to 0, \qquad R < I(X;Y) \text{ اگر .}$$

تا این جا فرض بر این بود که گیرنده و فرستنده W را به اشتراک گذاشته‌اند. با این کار در واقع تصادفی بودن کد را منطقی جلوه داده‌ایم، در صورتی که در عمل، گیرنده و فرستنده لزوماً چنین دنباله‌ای را در اختیار ندارند. برای حذف W می توان نوشت:

$$Pe_{ave} = \sum_w P(W = w)P(\text{خطا}|W = w).$$

در رابطه فوق، $P(W = w)$ احتمال اتفاق افتادن یک کتاب کد خاص را نشان می‌دهد. حال از لم زیر در ادامه استفاده می‌کنیم که اثبات آن به خواننده واگذار می‌شود.

لم ۱۰ اگر میانگین وزن‌دار یک سری عدد به سمت صفر میل کند، حتماً یکی از این اعداد به سمت صفر میل می‌کند.

از آنجایی که Pe_{ave} میانگین وزن‌دار یک سری احتمال است که به سمت صفر میل می‌کند، پس:

$$\exists w : P(\text{خطا}|W = w) \leq Pe_{ave} \to 0.$$

۴.۶ قسمت وارون

در قسمت وارون، می‌خواهیم ثابت کنیم که بیشتر از نرخ $\max_{p(x)} I(X;Y)$ نمی‌توانیم اطلاعات انتقال دهیم (با احتمال خطایی که به سمت صفر میل کند).

۱.۴.۶ روش اول

فرض کنید با نرخ R با n بار استفاده از کانال داده منتقل کرده‌ایم. داریم:

$$\begin{aligned}
nR = H(M) &\approx I(M;\hat{M}) && (۳)\\
&\leq I(X^n;Y^n) && (۴)\\
&= H(Y^n) - H(Y^n|X^n)\\
&\leq \sum_{i=1}^{n} H(Y_i|Y_1,Y_2,\cdots,Y_{i-1}) - \sum_{i=1}^{n} H(Y_i|Y_1,Y_2,\cdots,Y_{i-1}X^n)\\
&\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) && (۵)\\
&= \sum_{i=1}^{n} I(X_i;Y_i)\\
&\leq n \max_{p(x)} I(X;Y).
\end{aligned}$$

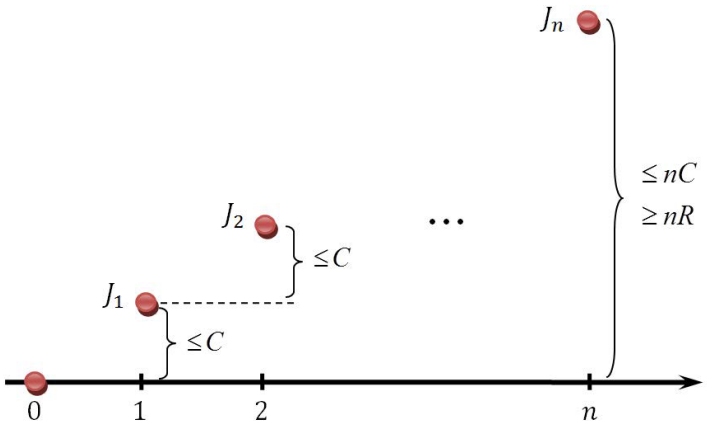
در روابط فوق، نامساوی (۳) بر اساس نامساوی فانو، و (۴) بر اساس اصل پردازش داده نوشته شده‌اند. در نامساوی (۴) نیاز به اثبات رابطه زیر داریم:

$$H(Y_i|Y_1,Y_2,\cdots,Y_{i-1}X^n) = H(Y_i|X_i).$$

این رابطه درست است زیرا با داشتن ورودی X_i به کانال i-ام تنها چیزی که مهم می‌باشد نویزی است که در این کانال اتفاق افتاده تا خروجی Y_i تولید شود. اما این نویز از نویزی که در کانال‌های قبلی اتفاق افتاده، و از محتویات کلمه کد ورودی مستقل است.

۲.۴.۶ روش دوم

فرض کنید مقدار nR بیت اطلاعات ارسال شده است. در گیرنده تمام n خروجی کانال را دریافت می‌کنیم و سپس عملیات کدبرداری را انجام می‌دهیم. اگر بتوانیم نشان دهیم که در هر مرحله استفاده از کانال بیشتر از C بیت اطلاعات نمی‌توانیم منتقل کنیم، مسئله اثبات شده است. لحظه i-ام را در نظر می‌گیریم. چون تمامی اطلاعات هنوز دریافت نشده است،



شکل ۸: نحوه تغییر میزان اطلاعات گیرنده از پیام ارسالی

نمیتوانیم کدبرداری را انجام دهیم. اما این نکته مشخص است که گیرنده و فرستنده در این لحظه چه چیزی دارند. یعنی محتوای اطلاعاتی فرستنده، MX^n و محتوای اطلاعاتی گیرنده، $Y_{1:i}$ است. میزان اطلاعات منتقل شده در مرحله i-ام را می‌توان از رابطه زیر حساب کرد:

$$J_i = I(MX^n; Y_{1:i}).$$

مشخص است که $J_0 = 0$. اما J_n بزرگتر یا مساوی nR است زیرا

$$J_n = I(MX^n; Y^n) = I(MX^n; Y^n\hat{M}) \geq I(M; \hat{M}) \approx H(M)$$

اگر بتوانیم نشان دهیم که $J_{i+1} - J_i \leq \max_{p(x)} I(X;Y)$ است، نشان داده‌ایم که $J_n \leq n \max_{p(x)} I(X;Y)$ زیرا

$$J_n = J_n - J_0 = (J_n - J_{n-1}) + (J_{n-1} - J_{n-2}) + \cdots + (J_2 - J_1) \leq n \max_{p(x)} I(X;Y)$$

در نتیجه

$$nR \leq J_n \leq n \max_{p(x)} I(X;Y)$$

که نتیجه مورد دلخواه ما را ثابت می‌کند.شکل ۸ موارد فوق را با یک گراف نمایش می‌دهد.

در ادامه نشان داده شده است که در هر مرحله بیشتر از $\max_{p(x)} I(X;Y)$ نمی‌توانیم اطلاعات منتقل کنیم:

$$\begin{aligned}
J_{i+1} - J_i &= I(MX^n; Y_1Y_2\cdots Y_{i+1}) - I(MX^n; Y_1Y_2\cdots Y_i) \\
&= I(MX^n; Y_{i+1}|Y_1Y_2\cdots Y_i) \\
&= H(Y_{i+1}|Y_1Y_2\cdots Y_i) - H(Y_{i+1}|Y_1Y_2\cdots Y_iMX^n) \\
&\leq H(Y_{i+1}) - H(Y_{i+1}|X_{i+1}) \\
&= I(Y_{i+1}; X_{i+1}) \\
&\leq \max_{p(x)} I(X;Y).
\end{aligned}$$

توجه کنید که مراحل اثبات روش دوم خیلی به مراحل اثبات روش اول شباهت دارد. اما تفسیری که از جملات نوشته شده بیان شده متفاوت است.