نظریه اطلاعات شبکه، ترم بهار ۱۳۹۴-۱۳۹۳

مدرس: امین زاده گوهری

# جلسه ۳ و ۴

## مرور بر مفاهیم اصلی نظریه اطلاعات نقطه به نقطه: کدگذاری منبع و کدگذاری کانال

---

## ۱ 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 ۱ 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.

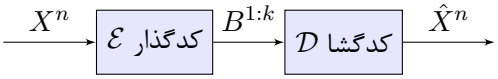
مفهوم بعدی که باید تعریف شود، احتمال خطای یک کد است. دقت کنید که احتمال خطای یک کد جزو پارامترهای تعریف آن نیست، بلکه چیزی است که پس از تعریف یک کد قابل محاسبه است.

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

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 $\epsilon$ داده شده (در اینجا همان مقدار $10^{-4}$) و $n \to \infty$، باید به کدهایی توجه شود که احتمال خطای آنها کمتر از $\epsilon$ و نرخ فشرده‌سازی آنها از همه کمتر باشد. به این نرخ کمینه $C_\epsilon$ می‌گوییم:

$$C_\epsilon := \inf_{\text{تمامی کدهایی که } P(\text{خطا}) \leq \epsilon} R.$$

در رابطه فوق $C_\epsilon$ کمترین فشرده‌سازی است به گونه‌ای که خطای بازیابی حداکثر $\epsilon$ باشد.




شکل ۱: نمایش شماتیک یک کدگذار منبع

در این صورت می‌توان ظرفیت فشرده‌سازی را این‌گونه تعریف کرد:

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

**قضیه ۱** در صورتی که یک منبع *i.i.d.* با توزیع $p(x)$ داشته باشیم ظرفیت فشرده‌سازی آن $C = H(X)$ است.

**تمرین ۲** نشان دهید که $C_\epsilon$ بر حسب $\epsilon$ تابعی غیرصعودی است. بعلاوه با فرض اینکه $p(x) > 0, \forall x \in \mathcal{X}$ ، نشان دهید که $C_0 = \log(|\mathcal{X}|)$. از اینجا نتیجه بگیرید که $C_\epsilon$ در صفر پیوسته نیست مگر اینکه توزیع $X$ یکنواخت باشد. همچنین مقدار $C_1$ را بیابید

**اثبات:**[ اثبات قضیه ۱] : جهت اثبات قابل حصول بودن میتوان از دو روش شماره گذاری دنباله های نوعی، و یا از روش سبدگذاری تصادفی استفاده کرد.

**حصول از طریق شماره گذاری دنباله های نوعی:** 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.

**حصول از طریق سبدگذاری تصادفی:** مجموعه تمامی دنباله های ممکن $x^n$ یعنی $\mathcal{X}^n$ را در نظر بگیرید. همچنین مجموعه $2^k$ سبد[۱] را در نظر بگیرید. برای ساختن کد هر کدام از اعضای $\mathcal{X}^n$ را بصورت تصادفی و یکنواخت در یکی از $2^k$ سبد قرار میدهیم. به این کار سبدگذاری تصادفی [۲] گفته میشود. فرض این است که نحوه سبدگذاری تصادفی بخشی از کد بوده و بر فرستنده و گیرنده آشکار است.[۳] عملیات کدگذاری به این شکل انجام میشود که فرستنده شماره سبد مربوط به دنباله ای که مشاهده کرده را برای گیرنده ارسال میکند. گیرنده تمامی دنباله هایی که در سبد مربوطه قرار دارند را در نظر میگیرد. تعداد این دنباله ها بطور متوسط $\frac{|\mathcal{X}|^n}{2^k}$ میباشد (زیرا $|\mathcal{X}|^n$ دنباله در $2^k$ سبد پرتاب شده اند). اما چون سبدگذاری کاملا بصورت تصادفی انجام شده بود، مجموعه دنباله هایی که در هر سبد قرار میگیرند کاملا تصادفی خواهد

[۱] Bin

[۲] Random Binning

[۳] میتوان این موضوع را اینگونه توجیه کرد که فرض کنیم که کدگذار و کدگشا یک رشته بیت تصادفی (مستقل از منبع $X^n$) را از قبل به اشتراک گذاشته اند. آنها از این بیت های تصادفی در انجام سبدگذاری تصادفی استفاده میکنند. پس از اینکه نشان دادیم که با استفاده از این منبع تصادفی به اشتراک گذاشته شده میتوان به احتمال خطای کوچک رسید، میتوانیم احتمال خطا را بصورت احتمال خطا به شرط دنباله های بیت های به اشتراک گذاشته نوشته و نتیجه بگیریم که حالت خاص و ثابتی از این بیت ها وجود دارد که احتمال خطا برای آن کوچک است.

بود. گیرنده به این دنباله ها در سبد مربوطه نگاه کرده و در صورتی که تنها یک دنباله نوعی در آن لیست باشد، آن را به عنوان دنباله ارسالی اعلام میکند. در غیر این صورت خطا اعلام میکند. بنابراین کافی است که خود را محدود به دنباله های نوعی و نحوه پخش شدن آنها در سبدها بکنیم و بقیه دنباله ها را در نظر نگیریم. ثابت میکنیم که احتمال کدگشایی درست در این روش به سمت یک میرود اگر $k > n(H(X) + \epsilon)$. در این صورت تعداد سبدها بصورت نمایی از تعداد دنباله های نوعی بیشتر خواهد بود و بصورت شهودی دنباله های نوعی با احتمال بالا در سبدهای متمایز قرار میگیرند. اما بصورت دقیق اگر یک دنباله نوعی خاص را در نظر بگیریم، احتمال خطا در کدگشایی آن برابر است با احتمال اینکه از میان $2^{nH(X)} - 1$ دنباله نوعی باقی مانده یکی در سبد مربوط به آن قرار بگیرید که با استفاده از باند مجموع [۴] این احتمال حداکثر $2^{nH(X)}2^{-k}$ است که در صورتی که $k > n(H(X) + \epsilon)$ با بزرگ شدن $n$ به سمت صفر میرود.

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

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

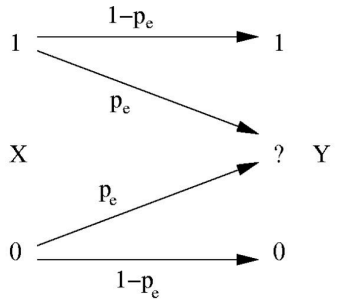
که در آن $B^{1:nR}$ همان دنباله فشرده شده از بیت‌ها است. □

# ۲ کدگذاری کانال

## ۱.۲ 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 ۳** 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:

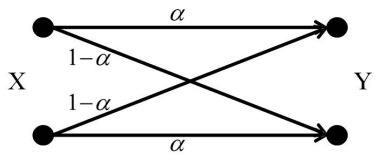
$$0\,1\,E\,1\,0\,\cdots\,0$$

[۴] Union Bound

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 ۴** 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.[۵] 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.

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

[۵] 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.

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

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:

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

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.

## ۲.۲ 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.[۶] 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 $\mathcal{Y} = \{0, E, 1\}$.

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


شکل ۲: نمایش شماتیک یک کدگذار کانال

**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 دنباله 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 |
|---|---|---|---|---|---|
| $1$ | $\cdots\ \cdots$ | $0$ | $1$ | $1$ | $1$ |
| $0$ | $\cdots\ \cdots$ | $0$ | $0$ | $1$ | $2$ |
| $1$ | $\cdots\ \cdots$ | $0$ | $1$ | $0$ | $3$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $0$ | $\cdots\ \cdots$ | $0$ | $1$ | $0$ | $2^{nR}$ |

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 $\epsilon$ داده شده (در اینجا همان مقدار $10^{-4}$) و $n \to \infty$، باید به کدهایی توجه شود که احتمال خطای آنها کمتر از $\epsilon$ و نرخ ارسال آنها از همه بیشتر باشد. به این نرخ ماکزیمم $C_\epsilon$ می‌گوییم:

$$C_\epsilon := \sup_{\substack{\text{تمامی کدهایی که } P(\text{خطا})\leq\epsilon}} R.$$

در رابطه فوق $C_\epsilon$ بیشترین نرخی است که می‌توان روی کانال ارسال کرد به گونه‌ای که خطا حداکثر $\epsilon$ 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.[۷]

**تمرین ۵** نشان دهید که $C_\epsilon$ بر حسب $\epsilon$ تابعی غیرنزولی است. بعلاوه با فرض اینکه $p(y|x) > 0, \forall x \in \mathcal{X}, y \in \mathcal{Y}$ ، نشان دهید که $C_0 = 0$. از اینجا نتیجه بگیرید که $C_\epsilon$ در صفر پیوسته نیست مگر اینکه ظرفیت شانون آن صفر باشد. همچنین مقدار $C_1$ را بیابید.

**تعریف ۶** به مقدار $C_0$ ظرفیت خطای صفر[۸] کانال میگویند که محاسبه آن به نظریه گراف ربط پیدا میکند و ریاضی دانان معروفی مانند لواژ[۹] روی آن کار کرده اند.

**مثال ۷** ظرفیت یک کانال $BSC$ را می‌توان به صورت زیر حساب کرد:

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

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

این یعنی اینکه جمله $H(Y|X)$ 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:

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

### ۱.۲.۲ ظرفیت حاصلضرب دو کانال

**تعریف ۸** برای دو کانال دلخواه $p_1(y_1|x_1)$ و $p_2(y_2|x_2)$ که روی الفبای $\mathcal{X}_1, \mathcal{Y}_1$ و $\mathcal{X}_2, \mathcal{Y}_2$ تعریف شده اند، کانال حاصل ضرب به این صورت تعریف میشود: الفبای ورودی کانال حاصل ضرب $\mathcal{X}_1 \times \mathcal{X}_2$ بوده، و الفبای خروجی آن $\mathcal{Y}_1 \times \mathcal{Y}_2$ میباشد. ضابطه کانال حاصل ضرب از روی حاصل ضرب ضابطه های تک تک کانال ها بدست می آید:

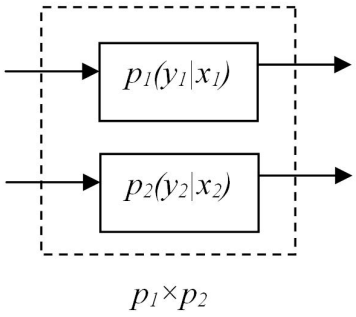
$$p(y_1, y_2|x_1, x_2) = p_1(y_1|x_1)p_2(y_2|x_2).$$

شکل ۳ ضرب دو کانال را نشان میدهد.

[۷] در بالا عبارت مربوط به ظرفیت را با ماکزیمم (و نه سوپریمم) نوشته شده است. در مورد کانال های با توزیع ورودی گسسته چون سیمپلکس احتمالاتی فضایی بسته و محدود با بعد محدود است، سوپریمم در واقع یک ماکزیمم است.

[۸] Zero-error capacity

[۹] Lovasz




شکل ۳: ضرب دو کانال $p_1$ و $p_2$.

ضرب دو کانال همانند این است که دو کانال را بصورت موازی و همزمان در نظر بگیریم و فرض کنیم که نویز هایی که این دو کانال را تحت تاثیر قرار میدهند از هم مستقل هستند. یعنی بصورت همزمان از دو کانال استفاده کنیم و بصورت همزمان دو سمبل بر روی دو کانال ارسال شود. در واقع ترکیب دو کانال را به عنوان یک کانال واحد در نظر بگیریم.

**تمرین ۹** توصیف حاصلضرب یک کانال $BSC$ و یک کانال $BEC$ را بصورت مشروح بنویسید.

حال قضیه زیر را در مورد کانال ضربی داریم:

**قضیه ۱۰** ظرفیت کانال حاصلضرب $p_1 \times p_2$ برابر جمع ظرفیت کانال های $p_1$ و $p_2$ است.

**اثبات:** در اینجا دو اثبات برای این موضوع ارائه میکنیم:

**اثبات اول:** داریم:

$$
\begin{aligned}
C_{p_1\times p_2} &= \max_{p(x_1,x_2)} I(X_1X_2;Y_1Y_2) && (۱)\\
&= \max_{p(x_1,x_2)} \big(I(X_1X_2;Y_1) + I(X_1X_2;Y_2|Y_1)\big)\\
&= \max_{p(x_1,x_2)} \big(I(X_1;Y_1) + I(X_2;Y_1|X_1) + I(X_2;Y_2|Y_1) + I(X_1;Y_2|X_2Y_1)\big)\\
&= \max_{p(x_1,x_2)} \big(I(X_1;Y_1) + I(X_2;Y_2|Y_1)\big) && (۲)\\
&\le \max_{p(x_1,x_2)} \big(I(X_1;Y_1) + I(Y_1X_2;Y_2)\big) && (۳)\\
&= \max_{p(x_1,x_2)} \big(I(X_1;Y_1) + I(X_2;Y_2) + I(Y_1;Y_2|X_2)\big)\\
&= \max_{p(x_1,x_2)} \big(I(X_1;Y_1) + I(X_2;Y_2)\big) && (۴)\\
&= \max_{p(x_1,x_2)} I(X_1;Y_1) + \max_{p(x_1,x_2)} I(X_2;Y_2)\\
&= \max_{p(x_1)} I(X_1;Y_1) + \max_{p(x_2)} I(X_2;Y_2)\\
&= C_{p_1} + C_{p_2},
\end{aligned}
$$

که در معادله (۱) با توجه به فرمول ظرفیت نوشته شده و بر اساس آن اجتماع روی تمامی توزیع های مشترک روی الفبای $\mathcal{X}_1 \times \mathcal{X}_2$ گرفته شده است، و نه تنها توزیع های مستقل. در معادله (۲) از رابطه زیر استفاده کردیم

$$I(X_2; Y_1|X_1) = I(X_1; Y_2|X_2Y_1) = 0$$

که با زنجیره مارکف های

$$X_2 - X_1 - Y_1, \qquad X_1 - X_2Y_1 - Y_2$$

معادل هستند. درستی این روابط را با کشیدن ساختار درختی مارکفی متغیرها میتوان مشاهده کرد. با داشتن ورودی کانال، خروجی کانال از بقیه متغیرها در زنجیره جدا خواهد شد و در نتیجه این روابط را داریم. معادله (۳) از این نظر اهمیت دارد که تکنیک کلی ای را در محاسبات روابط آنتروپیک بیان میکند. در اینجا متغیر $Y_1$ در قسمت شرطی آمده و میخواهیم آن را از قسمت شرطی حذف کنیم. اما میدانیم که حذف متغیر از قسمت شرطی ممکن است اطلاعات متقابل را زیاد یا کم کند. به این دلیل $Y_1$ را ابتدا از قسمت شرطی بیرون کشیده و در کنال $X_2$ قرار داده ایم. دلیل قرار دادن آن در کنار $X_2$ (و نه $Y_2$) از روابط بعدی مشخص میشود. نهایتا در معادله (۴) از زنجیره مارکف مشابه معادله (۲) استفاده کرده ایم.

پس روابط بالا نشان میدهد که $C_{p_1 \times p_2} \leq C_{p_1} + C_{p_2}$. اما از طرف دیگر

$$\begin{aligned}
C_{p_1 \times p_2} &= \max_{p(x_1,x_2)} I(X_1X_2; Y_1Y_2) \\
&\geq \max_{p(x_1)p(x_2)} I(X_1X_2; Y_1Y_2) \\
&= \max_{p(x_1)p(x_2)} \big(I(X_1; Y_1) + I(X_2; Y_2)\big) \\
&= \max_{p(x_1)} I(X_1; Y_1) + \max_{p(x_2)} I(X_2; Y_2) \\
&= C_{p_1} + C_{p_2},
\end{aligned}$$

که در روابط بالا از این نکته کلیدی استفاده شد که ماکزیمم گیری با اجتماع گیری روی تمام توزیع ها بیشتر یا مساوی ماکزیمم گیری با اجتماع روی توزیع های مستقل میباشد. این موضوع اثبات را کامل میکند.

**اثبات دوم:** از آنجایی که تابع اطلاعات متقابل نسبت به توزیع ورودی تابعی مقعر است، مساله یافتن

$$\max_{p(x_1,x_2)} I(X_1X_2; Y_1Y_2)$$

در زمره مسائل بهینه سازی محدب قرار میگیرد و مساله خوش دستی است زیرا هر ماکزیمم موضعی[۱۰] یک تابع مقعر یک ماکزیمم سراسری[۱۱] نیز هست. جهت چک کردن اینکه آیا یک نقطه ماکزیمم موضعی هست یا نه، کافی است که به مشتق های پاره آن نگاه کنیم: اگر نقطه کاملا درون ناحیه باشد، باید تمام مشتق های اول صفر باشد، و اگر نقطه در مرز ناحیه باشد، باید بررسی کنیم که مشتق در هنگام دور شدن از مرز بگونه ای باشد که تابع کاهش یابد. این شروط تحت

[۱۰] Local Maximum
[۱۱] Global Maximum

عنوان شرایط $KKT$ و با روش ضرایب لاگرانژ بدست می آیند. درستی آنها در مورد یک نقطه نتیجه میدهد که این نقطه ماکزیمم سراسری است.

فرض کنید که ماکزیمم $\max_{p(x_1)} I(X_1;Y_1)$ در نقطه $p^*(x_1)$ و $\max_{p(x_2)} I(X_2;Y_2)$ در نقطه $p^*(x_2)$ رخ بدهد. میخواهیم نشان دهیم که ماکزیمم $\max_{p(x_1,x_2)} I(X_1X_2;Y_1Y_2)$ در توزیع $p^*(x_1)p^*(x_2)$ رخ میدهد. کافی است که شرایط $KKT$ را برای این نقطه تحقیق کنیم.

برای مساله ماکزیمم کردن اطلاعات متقابل $\max_{p(x)} I(X;Y)$ شرایط $KKT$ به چه شکل بدست می آید. فرض کنید که توزیع $p^*(x)$ داده شده باشد. ابتدا برای هر $x_0 \in \mathcal{X}$ تعریف کنید:

$$I(x_0;Y) := \sum_y p(y|x_0) \log \frac{p^*(y,x_0)}{p^*(y)p^*(x_0)}.$$

که در آن $p^*(y,x_0) = p^*(x_0)p(y|x_0)$ و $p^*(y)$ با استفاده از توزیع ورودی $p^*(x)$ بدست می آید.[۱۲] شرایط $KKT$ به شرح زیر است: باید ثابت $\lambda \geq 0$ وجود داشته باشد بطوریکه

$$I(x_0;Y) = \lambda \qquad \forall x_0 : p^*(x_0) > 0,$$

$$I(x_0;Y) \leq \lambda \qquad \forall x_0 : p^*(x_0) = 0.$$

بازگردیم به مساله اصلی: میخواهیم نشان دهیم که ماکزیمم $\max_{p(x_1,x_2)} I(X_1X_2;Y_1Y_2)$ در توزیع $p^*(x_1)p^*(x_2)$ رخ میدهد. شرایط $KKT$ برای این نقطه به شرح زیر است:

$$I(x_1x_2;Y_1Y_2) = \lambda \qquad \forall x_1,x_2 : p^*(x_1)p^*(x_2) > 0,$$

$$I(x_1x_2;Y_1Y_2) \leq \lambda \qquad \forall x_1,x_2 : p^*(x_1)p^*(x_2) = 0.$$

از آنجایی که $p^*(x_1,x_2) = p^*(x_1)p^*(x_2)$ و $p(y_1,y_2|x_1,x_2) = p(y_1|x_1)p(y_2|x_2)$ میتوان نتیجه گرفت که $p^*(y_1,y_2) = p^*(y_1)p^*(y_2)$. در نتیجه

$$\begin{aligned}
I(x_1x_2;Y_1Y_2) &= \sum_{y_1y_2} p(y_1y_2|x_1x_2) \log \frac{p^*(y_1,y_2,x_1,x_2)}{p^*(y_1,y_2)p^*(x_1,x_2)} \\
&= \sum_{y_1y_2} p(y_1y_2|x_1x_2) \log \frac{p^*(y_1,x_1)p^*(y_2,x_2)}{p^*(y_1)p^*(y_2)p^*(x_1)p^*(x_2)}. \\
&= \sum_{y_1y_2} p(y_1y_2|x_1x_2) \big[ \log \frac{p^*(y_1,x_1)}{p^*(y_1)p^*(x_1)} + \log \frac{p^*(y_2,x_2)}{p^*(y_2)p^*(x_2)} \big] \\
&= \sum_{y_1y_2} p(y_1|x_1)p(y_2|x_2) \big[ \log \frac{p^*(y_1,x_1)}{p^*(y_1)p^*(x_1)} + \log \frac{p^*(y_2,x_2)}{p^*(y_2)p^*(x_2)} \big]. \\
&= I(x_1;Y_1) + I(x_2;Y_2).
\end{aligned}$$

[۱۲] با استفاده از تعریف داده شده میتوان نشان داد که $I(X;Y) = \sum_x p(x)I(x;Y)$. این رابطه انگیزه نامگذاری $I(x_0;Y)$ را بیان میکند

حال اگر $\lambda = \lambda_1 + \lambda_2$ مربوط به شرایط $KKT$ در زیرمساله های بهینه کردن اطلاعات متقابل قرار دهیم داریم: آنوقت شرط $p^*(x_1)p^*(x_2) > 0$ نتیجه میدهد که $p^*(x_1) > 0, p^*(x_2) > 0$ و در نتیجه

$$I(x_1; Y_1) = \lambda_1, I(x_2; Y_2) = \lambda_2.$$

پس $I(x_1x_2; Y_1Y_2) = \lambda_1 + \lambda_2 = \lambda$. اثبات حالت دیگر هم مشابه است. □

**نکته ۱۱** از روابط داده شده نتیجه میشود که جهت محاسبه $\max_{p(x_1,x_2)} I(X_1X_2; Y_1Y_2)$ کافی است که ماکزیمم گیری را روی توزیع های مستقل انجام دهیم. در بسیاری از مسائل نظریه اطلاعات با مساله مشابهی درگیر هستیم. یک ماکزیمم گیری روی یک فضا داریم و میخواهیم که محدوده ماکزیمم گیری را محدود کنیم. چنین محدود کردنی هم از لحاظ نظری و هم از لحاظ شبیه سازی کامپیوتری اهمیت دارد و میتوان حجم محاسبات و زمانی که باید برای یافتن نقطه ماکزیمم صرف کنیم را بشدت کاهش دهد.

**نکته ۱۲** نتیجه اثبات شده به این معنی است که اگر دو کانال موازی با نویزهای مستقل داشته باشیم. بهینه ترین کار تقسیم پیام به دو بخش و ارسال یک بخش روی کانال اول و بخش دوم روی کانال دوم است. کدگذاری مشترک روی این دو کانال ما را به نرخ بیشتری نمیرساند. یک کد به طول $n$ برای کانال ضربی $p_1 \times p_2$ از کانال ضربی $n$ بار استفاده میکند. پس از هر کدام از کانال های $p_1$ و $p_2$ به تعداد $n$ بار استفاده میشود. کدگذار پیام ورودی را دریافت و در ازای آن ورودی های کانال ها (کلمه کد) را مشخص میکند. بخشی از کلمه کد از روی نسخه های کانال اول و بخش دیگر از روی نسخه های کانال دوم عبور میکند. در گیرنده کدگشایی پس از دریافت خروجی های هر دو نسخه از کانال انجام میشود. نتیجه قضیه بیان شده این است که گیرنده کدگشایی را نه بصورت همزمان، بلکه بصورت جدا جدا روی نسخه های کانال $p_1$ و نسخه های کانال $p_2$ میتواند انجام دهد و این موضوع ضربه ای به گیرنده از لحاظ نرخ قابل حصول نمیرساند.

**تمرین ۱۳** کانالی با الفبای ورودی حرفی $\mathcal{X} = \{1, 2, 3, \cdots, n\}$ در نظر بگیرید بطوریکه برای هر سمبل خروجی $y$ رابطه زیر برقرار باشد:

$$p(y|x = n) = \sum_{i=1}^{n-1} \alpha_i p(y|x = i),$$

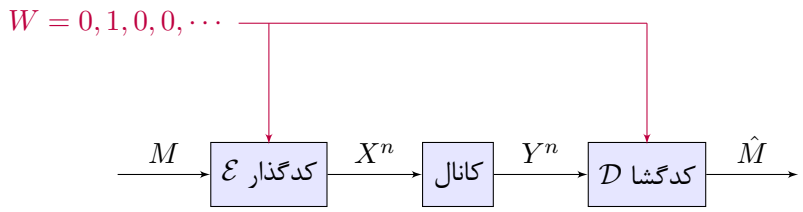
که در آن $\alpha_i$ ها اعدادی نامنفی با مجموعی یک هستند. با استفاده از روش شرایط کافی $KKT$ ثابت کنید که ظرفیت این کانال توسط توزیع احتمالی که مقدار آن روی سمبل $n$ ام صفر میباشد بدست می آید.

## ۳ قسمت مستقیم کانال نقطه به نقطه

منظور از قسمت مستقیم یا قسمت قابل حصول[۱۳] اثبات این است که نشان دهیم

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

[۱۳] Achievability part or the Direct part




شکل ۴: نمایش شماتیک یک کدگذار کانال به همراه منبع تصادفی به اشتراک گذاشته شده

ایده اثبات قسمت مستقیم رابطه ظرفیت، ایده کدگذاری تصادفی[۱۴] 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 ۴). 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.

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

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 این لیست شامل بیشتر از یک دنباله است. برای محاسبه احتمال خطا، فرض می‌کنیم پیام $M = 1$ ارسال شده است (به

[۱۴] Random Coding

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 اینکه "هیچ کلمه کدی با $Y^n$ 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:

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

### ۲.۰.۳ حذف منبع تصادفی به اشتراک گذاشته شده

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 ۱۴** 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:

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

### ۳.۰.۳ احتمال خطای بیشینه و متوسط

تا اینجا نشان دادیم که کدی وجود دارد که در آن احتمال خطای متوسط به سمت صفر میرود. احتمال خطای متوسط برابر است با:

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

احتمال خطای بیشینه به صورت زیر تعریف میشود:

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

اگر یک کد خوب داشته باشید (یعنی کدی که احتمال خطای متوسط کوچکی دارد)، می توان آن را هرس کرد و احتمال خطای بیشینه آن را کم کرد (هرس کردن یعنی کلمات کدی که مقدار $P(\text{خطا}|M = m)$ متناظرشان بزرگ است، را کنار می گذاریم). دقت کنید که کنار گذاشتن برخی از کلمات کد ممکن است نرخ کد را کم کند، اما احتمال خطای هیچکدام از کلمات کد دیگر را زیاد نمیکند (و ممکن است کم هم بکند).

مجموعه کلمات کد خوب را اینگونه تعریف میکنیم:

$$\{m|1 \leq m \leq 2^{nR}, \quad P(\text{خطا}|M = m) \leq 2\epsilon\}.$$

میخواهیم نشان دهیم که تعداد کلمات کد خوب زیاد است. برای این کار از نامساوی مارکف استفاده میکنیم:

**لم ۱۵** نامساوی مارکوف: اگر میانگین $k$ عدد کمتر یا مساوی $\epsilon$ باشد، حداقل نیمی از آن ها کمتر یا مساوی $2\epsilon$ هستند.

بنابراین با استفاده از این نامساوی داریم:

$$\left|\{m|1 \leq m \leq 2^{nR}, \quad P(\text{خطا}|M = m) \leq 2\epsilon\}\right| \geq \frac{2^{nR}}{2} = 2^{nR-1} = 2^{n(R-\frac{1}{n})}.$$

با اینکه تعداد کلمات کد کم شده است و در نتیجه نرخ کد از $R$ به $R-\frac{1}{n}$ کاهش می یابداما اگر $\lim_{n\to\infty} R-\frac{1}{n} = R$. در نتیجه میزان کم شدن نرخ قابل حصول با اضافه شدن طول کد کمتر و کمتر میشود.

## ۴ کدهای خطی تصادفی

روش حصول بیان شده برای رسیدن به ظرفیت در حالت کانالی با ورودی دودویی را بیاد آورید. کدگذار و کدبردار با استفاده از یک رشته تصادفی مشترک، کتاب کد را تولید می کنند. اما کد تولیدی لزوما کد خطی نیست. از آنجایی که کدهای خطی دارای ساختار بوده و در نظریه کدینگ دارای اهمیت میباشند، این سؤال پیش می آید که آیا با استفاده از کدهای خطی میتوان به ظرفیت شانون کانال رسید یا نه. البته برای اینکه این سؤال معنی دار باشد، باید فرض کنیم که

الفبای ورودی کانال یک میدان متناهی است. در صورتی که توزیع ورودی یکنواخت به ظرفیت برسد، جواب مثبت است. تحت این شرایط ثابت میکنیم که کدهای خطی تصادفی به ظرفیت شانون میرسند. برای مشخص کردن یک کد خطی باید ماتریس مولد با ابعاد $G_{n\times nR}$ را مشخص کنیم. فرض کنید که درایه های این ماتریس $G$ را بصورت تصادفی و $i.i.d$ از توزیع یکنواخت پر کنیم (با استفاده از رشته بیت های تصادفی مشترک). با استفاده از این ماتریس مولد میتوان تمامی $2^{nR}$ کلمه کد را لیست کرد. از آنجایی که ماتریس مولد تصادفی است، کلمات کد نیز تصادفی خواهند بود، اما قطعا کلمات کد کاملا از هم مستقل نیستند. باید به اثبات ارائه شده قبلی بازگردیم و ببینیم که در کجا و از چه چیزی استفاده کردیم. با بررسی اثبات خواهیم دید که کافیست که کلمات کد دوبدو مستقل از هم باشند ( نیازی به اینکه همه مستقل از هم باشند نیست). با فرض اینکه کلمه کد اول انتخاب شده است برای محاسبه احتمال خطا یک سری واقعه به صورت زیر تعریف کردیم:

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

سپس با استفاده از باند مجموع نوشتیم:

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

دقت کنید که جهت اثبات اینکه تمامی احتمالات $P(A_i)$ بالا کوچک هستند، نیاز به این داریم که کلمات کد مربوطه شان از کلمه کد ۱ بصورت مستقل انتخاب شده باشند. مثلا اگر کلمه کد ۲ از کلمه کد ۱ بصورت مستقل انتخاب شده باشد میتوان با همان استدلال نشان داد که $P(A_2)$ در حدود $2^{-nI(X;Y)}$ است. پس بدلیل استفاده از باند مجموع، تنها چیزی که به آن نیاز داریم استقلال دو به دوی کلمات کد است. نشان میدهیم که استقلال دو به دوی کلمات کد برقرار خواهد بود اگر درایه های ماتریس مولد را بصورت تصادفی و مستقل انتخاب کرده باشیم.

**قضیه ۱۶** اگر درایه های$G$ به صورت تصادفی از توزیع یکنواخت روی اعضای یک میدان انتخاب شوند، ثابت کنید برای هر دو بردار $n$ تایی دلخواه متفاوت $v_1$ و $v_2$ متغیر $Gv_2$ مستقل از $Gv_1$ است. دقت کنید که $Gv_2$ و $Gv_1$ متناظر دو سطر از جدول کتاب کد هستند.

**تمرین ۱۷** قضیه فوق را برای حالت خاصی که ورودی کانال $\mathcal{X}=\{0,1\}$ است ثابت کنید.

## ۵ قسمت وارون کانال نقطه به نقطه

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). در اینجا سه روش برای قسمت وارون ارائه میکنیم. روش اول روش استاندارد معمول

است که در آن رشته ای از نامساوی ها پشت سر هم می آید. هدف از اثبات روش دوم و سوم بیان شهودی تر نحوه نوشته شدن این رشته نامساوی ها در روش اول است. خصوصا روش سوم به ما یکی از جملاتی که در میانه رشته ظاهر میشود را مشخص کرده و با دانستن این جمله میانی به هم متصل کردن رشته نامساوی ها عملیاتی ساده تر خواهد بود. در طول این درس مکررا از روش سوم برای نوشتن اثبات های وارون استفاده خواهیم کرد.

### ۴.۰.۵ روش اول

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

$$\begin{aligned} nR = H(M) &\approx I(M;\hat{M}) && (۵)\\ &\leq I(X^n;Y^n) && (۶)\\ &= H(Y^n) - H(Y^n|X^n)\\ &\leq \sum_{i=1}^{n} H(Y_i|Y_1,Y_2,\cdots,Y_{i-1}) - \sum_{i=1}^{n} H(Y_i|Y_1,Y_2,\cdots,Y_{i-1}X^n)\\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) && (۷)\\ &= \sum_{i=1}^{n} I(X_i;Y_i)\\ &\leq n \max_{p(x)} I(X;Y). \end{aligned}$$

در روابط فوق، نامساوی (۵) بر اساس نامساوی فانو، و (۶) بر اساس اصل پردازش داده نوشته شده‌اند. در نامساوی (۶) نیاز به اثبات رابطه زیر داریم:

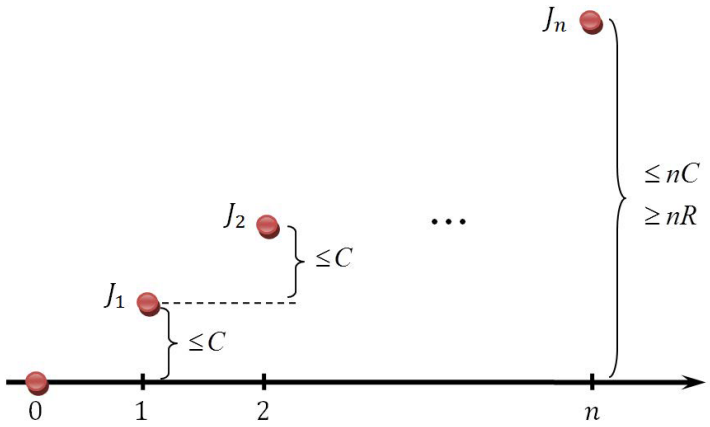
$$H(Y_i|Y_1,Y_2,\cdots,Y_{i-1}X^n) = H(Y_i|X_i).$$

این رابطه درست است و در بحث مربوط به زنجیره مارکف به آن پرداختیم. 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.

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

$$J_i = I(M;Y_{1:i}).$$




شکل ۵: نحوه تغییر میزان اطلاعات گیرنده از پیام ارسالی

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

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

توجه کنید که این بخش از نامساوی ها شبیه بخش اول رشته نامساوی ها در اثبات اول است. اگر بتوانیم نشان دهیم که $J_{i+1} - J_i \leq \max_{p(x)} I(X;Y)$ زیرا $J_n \leq n \cdot \max_{p(x)} I(X;Y)$ است، نشان داده‌ایم که

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

در نتیجه

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

که نتیجه مورد دلخواه ما را ثابت می‌کند. شکل ۵ 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:

$$\begin{aligned}
J_{i+1} - J_i &= I(M; Y_1 Y_2 \cdots Y_{i+1}) - I(M; Y_1 Y_2 \cdots Y_i) \\
&= I(M; Y_{i+1} | Y_1 Y_2 \cdots Y_i) \\
&= H(Y_{i+1} | Y_1 Y_2 \cdots Y_i) - H(Y_{i+1} | Y_1 Y_2 \cdots Y_i M) \\
&= H(Y_{i+1} | Y_1 Y_2 \cdots Y_i) - H(Y_{i+1} | Y_1 Y_2 \cdots Y_i M X^n) \\
&\leq H(Y_{i+1}) - H(Y_{i+1} | X_{i+1}) \\
&= I(Y_{i+1}; X_{i+1}) \\
&\leq \max_{p(x)} I(X;Y).
\end{aligned}$$

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. در واقع تنها نامساوی ای که در اینجا زده ایم

$$H(Y_{i+1}|Y_1Y_2\cdots Y_i) \leq H(Y_{i+1})$$

میباشد. در اثبات قبلی هم دقیقا همین نامساوی ها را زده بودیم:

$$H(Y^n) = \sum_i H(Y_i|Y_{1:i-1}) \leq \sum_i H(Y_i).$$

همچنین در بالا از زنجیره مارکف

$$Y_{i+1} - X_{i+1} - Y_1Y_2\cdots Y_iM$$

استفاده کردیم که مشابه زنجیره مارکف استفاده شده در روش قبلی میباشد.

## ۱.۵ روش سوم

### ۱.۱.۵ مقدمات: ظرفیت به عنوان تابعی از کانال

اگر یک کانال تغییر کند، ظرفیت آن هم تغییر می کند، پس ظرفیت هم تابعی از کانال است.

$$C\big(q(y|x)\big) = \max_{p(x)} I(X;Y).$$

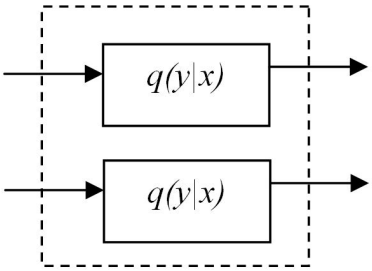
ما تا اینجا کانال را ثابت فرض کرده بودیم و این وابستگی فرمول $C$ به کانال را بصورت مشخص نمی نوشتیم. ظرفیت به عنوان تابعی از کانال دارای ویژگی هایی است. مثلا در رابطه بالا چون عبارت داده شده جبری است ظرفیت بر حسب کانال تابعی پیوسته خواهد بود.

**مثال ۱۸** ثابت میکنیم که $C\big(q(y|x)\big) = \max_{p(x)} I(X;Y)$ تابعی محدب از $q(y|x)$ است. توصیه میکنیم که قبلا از خواندن اثبات سعی کنید خودتان آن را ثابت کنید. فرض کنید که $q_1(y|x)$ و $q_2(y|x)$ دو کانال داده شده باشند. تعریف کنید $q(y|x) = \frac{1}{2}q_1(y|x) + \frac{1}{2}q_2(y|x)$. میدانیم که به ازای یک $p(x)$ ثابت $I(X;Y)$ تابعی محدب در کانال $p(y|x)$ میباشد. در نتیجه

$$I_q(X;Y) \leq \frac{1}{2}\big(I_{q_1}(X;Y) + I_{q_2}(X;Y)\big), \qquad \forall p(x).$$

پس

$$\begin{aligned} C(q) = \max_{p(x)} I_q(X;Y) &\leq \frac{1}{2}\max_{p(x)}\big(I_{q_1}(X;Y) + I_{q_2}(X;Y)\big) \\ &\leq \frac{1}{2}\big(\max_{p(x)} I_{q_1}(X;Y) + \max_{p(x)} I_{q_2}(X;Y)\big) \\ &= \frac{1}{2}(C(q_1) + C(q_2)). \end{aligned}$$




شکل ۶: ضرب دو کانال یکسان $q$

**نکته ۱۹** یکی از نکات اصلی یک محقق خوب بودن این است که پس از خواندن هر اثباتی از خود بپرسید که این اثبات دقیقا از چه فرض هایی استفاده کرده و یا اینکه آیا میتوان قضیه ای کلی تر را با استفاده از ایده های مطرح شده در آن بیان کرد. مثلا در مورد اثبات بالا میتوان لم زیر را فرمول بندی کرد که به نظر می آید نکته اصلی اثبات بوده است، و به خاطر سپردن آن (یا ایده اصلی اثبات آن) میتواند در مسائلی که در آینده به آن برخورد میکنید مفید باشد. لم: تابع دلخواه تابع $g(x,y)$ را در نظر بگیرید بطوریکه برای هر $x$ ثابت، $g(x,y)$ نسبت به $y$ محدب باشد. در این صورت $f(y)=\max_x g(x,y)$ نیز نسبت به $y$ تابعی محدب است.

مشابه ناحیه ظرفیت، وقتی صحبت از کران بالای ظرفیت $T$ میکنیم، باز هم صحبت از $T\big(q(y|x)\big)$ که تابعی از کانال است میکنیم.

**تعریف ۲۰** $T\big(q(y|x)\big)$ یک کران بالا برای ظرفیت است اگر و فقط اگر

$$T\big(q(y|x)\big) \geq C\big(q(y|x)\big), \qquad \forall q(y|x).$$

### ۲.۱.۵ مقدمات: خواص عملیاتی ظرفیت

ادعا می کنیم که برای هر کانال دلخواه $q(y,z|x)$ داریم:

$$\frac{1}{2}C(q\times q)=C(q).$$

شکل ۶ ضرب دو کانال یکسان $q$ را را نشان میدهد. اثبات این رابطه با توجه به تعریف عملیاتی از ظرفیت قابل اثبات است. برای اثبات یک کد برای کانال $q\times q$ در نظر بگیرید. اگر از این کانال، $n$ بار استفاده شود و در کل $nR$ بیت از کانال ارسال شود، معادل آن است که از کانال $q$ به میزان $2n$ بار استفاده شده است. در نتیجه این کد را میتوان یک کد برای انتقال اطلاعات روی کانال $q$ قلمداد کرد. و در این صورت نرخ ارسال برای کانال $q$ برابر است با

$$\frac{nR}{2n}=\frac{R}{2}.$$

برعکس یک کد برای کانال $q$ در نظر گرفته بگیرید. اگر از این کانال، $n$ بار استفاده شود و در کل $nR$ بیت از کانال ارسال شود، معادل آن است که از کانال $q\times q$ به میزان $\frac{1}{2}n$ بار استفاده شده است. در نتیجه این کد را میتوان یک کد برای

انتقال اطلاعات روی کانال $q \times q$ قلمداد کرد که به نرخ ارسال $2R$ روی کانال ضربی میرسد. از آنجایی که تعریف عملیاتی ظرفیت ماکزیمم نرخ قابل حصول بشکل قابل اطمینان است، رابطه مورد نظر ثابت شده است.

بصورت مشابه در حالت کلی داریم:

$$\frac{1}{n}C(\underbrace{q \times q \times q \cdots q}_{n}) = C(q), \qquad \forall n.$$

اثبات مشابه حالت خاص $n = 2$ میباشد.

## ۲.۵ نگاه سوم به وارون نقطه به نقطه

قرار دهید

$$T(q(y|x)) = \max_{p(x)} I(X;Y).$$

هدف ما ثابت کردن این است که $T \geq C$. اما پیش از آن برای یک لحظه فرض کنید که میدانیم که $T = C$. با این فرض بایستی انتظار داشته باشیم که همان روابطی که برای $C$ درست هستند برای $T$ هم درست باشند. پس بایستی انتظار داشته باشیم که روابط زیر برقرار باشد:

$$\frac{1}{n}T(\underbrace{q \times q \times q \cdots q}_{n}) = T(q), \qquad \forall n \tag{۸}$$

$$\max_{n} \frac{1}{n}T(\underbrace{q \times q \times q \cdots q}_{n}) \geq C(q) \tag{۹}$$

رابطه (۹) باید برقرار باشد زیرا سمت چپ رابطه همان $T(q)$ است که با $C(q)$ مساوی است. اما ما آن را بصورت نامساوی نوشته ایم که دلیل آن در زیر مشخص میشود.

**نکته ۲۱** عبارت $\frac{1}{n}T(\underbrace{q \times q \times q \cdots q}_{n})$ به شکل $n$ حرفی $T(q)$ معروف است.

حال برگردیم به مساله اصلی که میخواهیم ثابت کنیم $T(q(y|x)) = \max_{p(x)} I(X;Y)$ کران بالایی برای ظرفیت است. برای این کار از قضیه زیر استفاده میکنیم که اثبات آن تقریبا واضح است.

**قضیه ۲۲** اگر تابع دلخواه $T(q)$ در معادلات (۸) و (۹) صدق کند، آنوقت

$$T(q(y|x)) \geq C(q(y|x)) \qquad \forall q(y|x).$$

خواهیم دید که ثابت کردن روابط (۸) و (۹) در واقع چیزی جز دنبال کردن همان اثبات های ذکر شده قبلی نیست. در واقع این روش تفسیری متفاوت از آنچه قبلا بیان شده بود بدست میدهد.

یک کانال خاص $q$ در نظر بگیرید. اول معادله (۹) بعد معادله (۸) را ثابت میکنیم. یک کد $(n, \epsilon)$ با نرخ نزدیک به ظرفیت انتخاب کنید که $n$ بار از کانال استفاده کند و به احتمال خطای $\epsilon$ برسد. فرض کنید که فاصله نرخ این کد از ظرفیت حداکثر $\delta$ باشد. متغیرهای مربوط به این روش کدگذاری را در نظر بگیرید:

$$M \to X^n \to Y^n \to \widehat{M}$$

$$P(M \neq \widehat{M}) \leq \epsilon, \qquad R \geq C(q) - \delta$$

حال $n$ مربوط به این کد را در نظر بگیرید. نشان میدهیم

$$\frac{1}{n} T(\underbrace{q \times q \times q \cdots q}_{n}) \geq C(q) - \delta. \tag{۱۰}$$

اثبات این رابطه فرمول (۹) را با میل دادن $\delta$ به سمت صفر ثابت میکند.

جهت اثبات (۱۰)، چند جمله اول اثبات قبلی را بیاد می آوریم. با استفاده از قضیه پردازش داده نوشته بودیم:

$$nR = H(M) \approx I(M; \widehat{M}) \leq I(X^n; Y^n) \tag{۱۱}$$

توجه کنید که

$$\frac{1}{n} T(\underbrace{q \times q \times q \cdots q}_{n}) = \frac{1}{n} \sup_{p(x_1, x_2, \cdots, x_n)} I(X^n; Y^n)$$

اما یک انتخاب خاص برای $p(x_1, x_2, \cdots, x_n)$ همان توزیعی است که توسط کد داده شده القا میشود. سپس با استفاده از معادله (۱۱) خواهیم داشت

$$\frac{1}{n} T(\underbrace{q \times q \times q \cdots q}_{n}) \geq R \geq C - \delta.$$

پس (۹) اثبات شد. در نتیجه اثبات فرمول (۹) چیزی جز چند خط اول اثبات قبلی نبود.

**نکته ۲۳** نگاه سوم به وارون نقطه به نقطه به ما میگوید که چرا در طول اثبات قبلی از جمله $I(X^n; Y^n)$ عبور کردیم. چون این جمله شکل $n$ حرفی مربوط به کرانی است که میخواهیم آن را ثابت کنیم . در تمامی وارون هایی که در آینده خواهیم دید به این نکته باز میگردیم. پس از چند مرحله استفاده از فانو و قضیه پردازش داده و غیره باید به عباراتی برسیم که شکل $n$ حرفی ناحیه ای که میخواهیم کران بیرونی بودن آن را ثابت کنیم باشد. پس طی چند مرحله ابتدایی هر اثبات وارون کاملا نظام مند خواهد بود، اگر حدس مناسبی از ناحیه کران بیرونی ای که آن را میخواهیم اثبات کنیم داشته باشیم.

حالا میخواهیم رابطه (۸) را ثابت کنیم. میبینیم که بقیه اثبات همانند اثبات قبلی است. در اینجا آن را برای $n = 2$ می نویسیم.

$$\begin{aligned}
\sup_{p(x_1,x_2)} \frac{1}{2} I(X_1X_2;Y_1Y_2) &= \sup_{p(x_1,x_2)} \frac{1}{2}\big(H(Y_1,Y_2) - H(Y_1,Y_2|X_1,X_2)\big) \\
&\leq \sup_{p(x_1,x_2)} \frac{1}{2}\big(H(Y_1) + H(Y_2) - H(Y_1|X_1,X_2) - H(Y_2|X_1,X_2,Y_1)\big) \\
&= \sup_{p(x_1,x_2)} \frac{1}{2}\big(H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2)\big) \\
&= \sup_{p(x_1,x_2)} \frac{1}{2}\big(I(X_1;Y_1) + I(X_2;Y_2)\big) \\
&= \frac{1}{2}\sup_{p(x_1)} I(X_1;Y_1) + \frac{1}{2}\sup_{p(x_2)} I(X_2;Y_2) \\
&= \sup_{p(x)} I(X;Y)
\end{aligned}$$

**تمرین ۲۴** اثبات بالا را با اثبات اول قضیه ۱۰ مقایسه کنید.

**مثال ۲۵** تحقیق مستقیم رابطه $\frac{1}{2}T(q \times q) = T(q)$ برای کانال $BEC$: فرض کنیم ضرب دو کانال $BEC$ با احتمال پاک شدن $e$ را داریم. باید ثابت کنیم

$$\sup_{p(x_1,x_2)} I(X_1,X_2;Y_1,Y_2) \leq 2\sup_{p(x)} I(X;Y).$$

جهت سادگی محاسبات به جای الفبای 0 و 1 برای ورودی کانال از 1 و $-1$، و همچنین به جای الفبای 0 و 1 و $E$ برای خروجی بترتیب از 1 و $-1$ و 0 استفاده می کنیم. در این صورت $Y_1$ را میتوان به صورت زیر نوشت: $Y_1 = X_1 \cdot T_1$ که در آن $T_1$ متغیری مستقل از $X_1$ و دارای توزیع زیر است:

$$T_1 = \begin{cases} 1, & \text{با احتمال } e \\ 0, & \text{با احتمال } 1-e \end{cases}$$

همچنین میتوان فرض کرد که در خروجی هم $T_1$ و $X_1 \cdot T_1$ دریافت میشود زیرا زمانی که خروجی پاک بود گیرنده آن را کشف میکند. این متغیرها محاسبات را آسان میکنند، مثلا

$$\begin{aligned}
I(X_1;Y_1) &= I(X_1;X_1 \cdot T_1, T_1) \\
&= I(X_1;T_1) + I(X_1;X_1 \cdot T_1|T_1) \\
&= I(X_1;X_1 \cdot T_1|T_1) \\
&= p(T_1=0) \times I(X_1;X_1 \cdot T_1|T_1=0) + p(T_1=1) \times I(X_1;X_1 \cdot T_1|T_1=1) \\
&= p(T_1=0) \times I(X_1;0|T_1=0) + p(T_1=1) \times I(X_1;X_1|T_1=1) \\
&= (1-e)H(X_1).
\end{aligned}$$

از این رابطه میتوان نتیجه گرفت که

$$\sup_{p(x_1)} I(X_1;Y_1) = \sup_{p(x_1)} (1-e)H(X_1) = 1-e.$$

سپس عبارت دو حرفی را محاسبه میکنیم:

$$\begin{aligned}\sup_{p(x_1,x_2)} I(X_1,X_2;Y_1,Y_2) &= \sup_{p(x_1,x_2)} I\big(X_1,X_2;T_1,T_2,(X_1\cdot T_1),(X_2\cdot T_2)\big)\\ &= \sup_{p(x_1,x_2)} I(X_1,X_2;T_1,T_2) + I(X_1,X_2;X_1\cdot T_1, X_2\cdot T_2|T_1,T_2)\\ &= \sup_{p(x_1,x_2)} \big[0 + I(X_1,X_2;X_1\cdot T_1, X_2\cdot T_2|T_1,T_2)\big].\end{aligned}$$

متغیرهای $T_1$ و $T_2$ مستقل هستند چون دو کانال موازی هستند. پس روابط را ادامه میدهیم:

$$\begin{aligned}\sup_{p(x_1,x_2)} I(X_1,X_2;Y_1,Y_2) &= \sup_{p(x_1,x_2)} I(X_1,X_2;X_1\cdot T_1, X_2\cdot T_2|T_1,T_2)\\ &= \sup_{p(x_1,x_2)} \big[e(1-e)I(X_1,X_2;X_1\cdot T_1, X_2\cdot T_2|T_1=0,T_2=1)\\ &\qquad + e(1-e)I(X_1,X_2;X_1\cdot T_1, X_2\cdot T_2|T_1=1,T_2=0)\\ &\qquad + e^2 I(X_1,X_2;X_1\cdot T_1, X_2\cdot T_2|T_1=0,T_2=0)\\ &\qquad + (1-e)^2 I(X_1,X_2;X_1\cdot T_1, X_2\cdot T_2|T_1=1,T_2=1)\big]\\ &= \sup_{p(x_1,x_2)} \big[e(1-e)H(X_1) + e(1-e)H(X_2) + (1-e)^2 H(X_1,X_2)\big]\\ &\le e(1-e) + e(1-e) + 2(1-e^2)\\ &= 2(1-e)\\ &= \sup_{p(x_1)} I(X_1;Y_1) + \sup_{p(x_2)} I(X_2;Y_2).\end{aligned}$$

## ۶ نکات تکمیلی در مورد ظرفیت کانال نقطه به نقطه

در بخش قبل سه اثبات برای وارون مساله نقطه به نقطه مطرح کردیم. روش های دیگری برای اثبات وارون هم وجود دارد، خصوصا اگر با کانال خاصی مواجه باشیم. مثلا کانال $BSC$ با پارامتر $p$ را در نظر بگیرید. این کانال را میتوان به این شکل مدل کرد: $Y = (X+Z) \mod 2$ که در آن $X \in \{0,1\}$ ورودی کانال و $Z \in \{0,1\}$ نویز کانال میباشد. در صورتی که گیرنده از روی $Y^n$ بتواند دنباله $X^n$ را بیابد، میتواند دنباله $Z^n$ را هم بیابد، چون $Z$ تابعی از ورودی و خروجی کانال است. پس اگر یک کانال مجازی از نویز محیط به گیرنده در نظر بگیریم، نرخ انتقال اطلاعات نویز $h(p)$ خواهد بود. چون گیرنده یک بیت دریافت میکند حداکثر به اندازه $1-h(p)$ بیت برای انتقال پیام باقی می ماند. پس $C \le 1-h(p)$. بصورت دقیق ریاضی داریم:

$$n \ge H(Y^n) \ge I(Y^n; MZ^n) \ge I(\hat{M}\hat{Z}^n; MZ^n) \approx H(M) + nH(Z) = nR + nh(p)$$

پس $R \leq 1 - h(p)$.

### ۱.۶ نامساوی های اثباتهای وارون

یک وارون خوب، وارونی است که از نامساوی هایی استفاده کند که بتوانند به تساوی تبدیل شوند، چه در غیر این صورت وارون ضعیفی خواهد بود. از آنجایی که وارون نقطه به نقطه ما محکم است (با قسمت قابل حصول آن جفت میشود)، پس باید انتظار داشته باشیم که نامساوی هایی که در آن استفاده شده برای هر کدی که به ظرفیت میرسد به تساوی تبدیل شوند. در قسمت وارون این نامساوی را داشتیم:

$$H(Y^n) \leq \sum_{i=1}^{n} H(Y_i).$$

تساوی زمانی برقرار میشود که $Y_i$ ها مستقل باشند. همچنین داشتیم:

$$\sum_{i=1}^{n} I(X_i; Y_i) \leq n \cdot \max_{p(x)} I(X; Y).$$

با فرض اینکه توزیعی که ظرفیت را ماکزیمم میکند یکتا و برابر $p^*(x)$ است نتیجه میگیریم که $p(x_i) \sim p^*(x_i)$. در نتیجه توزیع ورودی های کانال یکسان است، و در نتیجه توزیع خروجی ها هم باید یکسان باشد. یعنی یک کد خوب در خروجی کانال بصورت تقریبی یک دنباله $i.i.d.$ القا میکند. اما یک دنباله $i.i.d.$ به معنی یک توزیع یکنواخت روی دنباله های نوعی خروجی می باشد. بعدها که در مورد لم گنجایشی و پوششی صحبت میکنیم به این موضوع در مورد کدهایی که بصورت تصادفی تولید میشوند، باز میگردیم.

## ۷ کانال های با شرط هزینه در ورودی

در برخی از موارد ممکن است که روی دنباله مجاز برای ورودی کانال شرطی وجود داشته باشد (مثلا در کانال های با الفبای ورودی گسسته درصد بارهایی که از یک سمبل خاص در ورودی استفاده میشود، و یا یک شرط توان در کانال های با الفبای ورودی پیوسته). مثلا فرض کنید که یک تابع هزینه $b(x)$ برای استفاده از سمبل ورودی $x$ وجود داشته باشد.[۱۵] دو گونه محدودیت هزینه متوسط حداکثر $B$ روی ورودی های کانال میتوان متصور شد (مشابه احتمال خطا که هم شرط مقدار متوسط خطا داشتیم و هم شرط مقدار بیشینه):[۱۶]

۱. محدودیت هزینه روی تمامی کلمات کد: فرض میکنیم که

$$\frac{1}{n} \sum_{i=1}^{n} b(x_i(m)) \leq B, \qquad \forall m.$$

که در آن $x_i(m)$ ورودی به کانال در لحظه $i$ ام است زمانی که پیام $m$ را ارسال میکنیم.

[۱۵] کتاب الجمال و کیم فرض میکند که شرط های اضافه $b(x) \geq 0 \ \forall x$ و وجود سمبل بدون هزینه $x_0$ که $b(x_0) = 0$ نیز در مورد تابع هزینه برقرار است. روش ما اندکی با روش بیان شده در کتاب متفاوت است و در نتیجه نیازی به این شرط ها نداریم. اما سادگی اثبات کتاب نیز آموزنده است. نقطه قوت اثبات ما نداشتن ایده جدید و موازی بودن با اثبات مربوط به احتمال متوسط خطا و احتمال بیشینه خطا است.

[۱۶] کتاب الجمال و کیم تنها اولین محدودیت را در نظر میگیرد.

۲. محدودیت هزینه متوسط روی کلمات کد: فرض میکنیم که

$$\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\frac{1}{n}\sum_{i=1}^{n}b(x_i(m))\leq B.$$

**مثال ۲۶** فرض کنید که الفبای ورودی دودویی است و مجاز باشیم که از ورودی 1 در حداکثر یک سوم زمان ها استفاده کنیم (یعنی از دنباله کدهایی که درصد یک ها در آنها حداکثر یک سوم باشد). در این صورت محدودیت هزینه روی تمامی کلمات کد داریم با $b(0)=0, b(1)=1$، $B=\frac{1}{3}$.

**قضیه ۲۷** ظرفیت یک کانال با محدودیت هزینه $B$ روی $X$ در دو حالت "هزینه متوسط" و "هزینه روی تمامی کلمات کد" مساوی بوده و برابر است با

$$C(B)=\max_{p(x):\mathbb{E}[b(X)]\leq B}I(X;Y). \qquad (۱۲)$$

پیش از اثبات این قضیه ثابت میکنیم که $C(B)$ بر حسب $B$ تابعی غیرنزولی، مقعر و پیوسته میباشد. غیرصعودی بودن تقریبا واضح است. اثبات پیوسته بودن $C(B)$ در $B$ در پیوست آمده است. برای اثبات مقعر بودن فرض کنید که ماکزیمم $C(B_1)$ در $p_1(x)$ و ماکزیمم $C(B_2)$ در $p_2(x)$ بدست می آید. در این صورت اگر $B=\frac{1}{2}(B_1+B_2)$ آنوقت تعریف کنید $p(x)=\frac{1}{2}(p_1(x)+p_2(x))$. در این صورت

$$\mathbb{E}_p[b(X)]=\frac{1}{2}\big(\mathbb{E}_{p_1}[b(X)]+\mathbb{E}_{p_2}[b(X)]\big)\leq\frac{1}{2}(B_1+B_2)=B.$$

و همچنین از مقعر بودن اطلاعات متقابل نسبت به توزیع ورودی داریم:

$$I_p(X;Y)\geq\frac{1}{2}\big(I_{p_1}(X;Y)+I_{p_2}(X;Y)\big).$$

در نتیجه

$$C(B)\geq I_p(X;Y)\geq\frac{1}{2}\big(I_{p_1}(X;Y)+I_{p_2}(X;Y)\big)=\frac{1}{2}\big(C(B_1)+C(B_2)\big).$$

اثبات قابل حصول بودن و وارون مشابه حالت بدون شرط هزینه ورودی است.

**اثبات وارون:** کافی است که نشان دهیم که نرخ بیش از آنچه در معادله (۱۲) آمده در حالت "هزینه متوسط" قابل حصول نیست (و در نتیجه در حالت "هزینه روی تمامی کلمات کد" نیز نمیتواند قابل حصول باشد). مشابه قبل داریم

$$R\leq\frac{1}{n}\sum_{i=1}^{n}I(X_i;Y_i)$$

تعریف کنید

$$B_i=\mathbb{E}[b(X_i)].$$

در نتیجه

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} B_i &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[b(X_i)] \\ &= \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} b(X_i)] \\ &= \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} b(X_i)|M=m] \\ &= \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\frac{1}{n}\sum_{i=1}^{n} b\big(x_i(m)\big) \\ &\leq B.\end{aligned}$$

با استفاده از تعریف $C(B)$ در معادله داریم:

$$I(X_i;Y_i) \leq C(B_i).$$

در نتیجه

$$\begin{aligned}R &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i) \leq \frac{1}{n}\sum_{i=1}^{n} C(B_i) \\ &\leq C(\frac{1}{n}B_i) \leq C(B).\end{aligned}$$

که در معادلات بالا از $C(B)$ بر حسب $B$ تابعی غیرنزولی و مقعر میباشد استفاده کرده ایم.

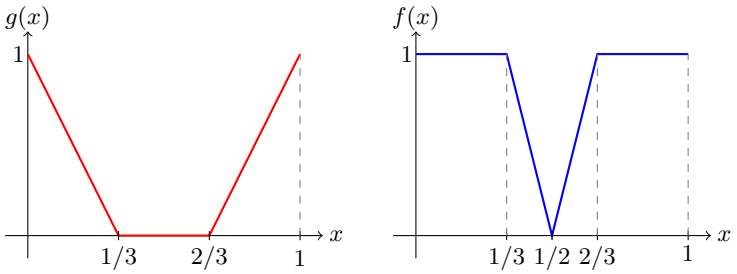
**اثبات قابل حصول:** ابتدا نشان میدهیم که نرخ معادله (۱۲) زمانی که شرط "هزینه متوسط" را داشته باشیم قابل حصول است. رسیدن به شرط "هزینه روی تمامی کلمات کد" کاملا مشابه روش رسیدن از مقدار متوسط خطا به مقدار بیشینه خطا از طریق هرس کردن کتاب کد است. یک $\epsilon$ ثابت در نظر بگیرید. اگر فرض کنیم که یک کتاب کد با اندازه $2^{nR}$ داریم که شرط "هزینه متوسط" $\frac{B}{1+\epsilon}$ در مورد آن برقرار باشد. آنوقت تمامی کلمات کدی که هزینه آنها کمتر مساوی از $B$ است را در نظر میگیریم. طبق نامساوی مارکف اینها حداقل

$$\frac{\epsilon}{1+\epsilon}2^{nR} = 2^{n(R+\frac{1}{n}\log\frac{\epsilon}{1+\epsilon})}$$

میباشد. اما برای $\epsilon$ ثابت، اگر $n$ به سمت بینهایت برود $\frac{1}{n}\log\frac{\epsilon}{1+\epsilon}$ به سمت صفر میرود و در نتیجه ما به همان نرخ $R$ میرسیم.

نشان میدهیم که نرخ معادله (۱۲) زمانی که شرط "هزینه متوسط" را داشته باشیم قابل حصول است. توزیع $p(x)$ را در نظر بگیرید که به $C(\frac{B}{1+\epsilon})$ میرسد. کتاب کد را بصورت $i.i.d.$ از این توزیع تولید میکنیم. در اینجا بجز واقعه خطاهایی که قبلا داشتیم، یک واقعه خطای جدید هم داریم و آن اینکه شرط هزینه برای کتاب کد تصادفی برقرار نباشد:

$$\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\frac{1}{n}\sum_{i=1}^{n} b\big(X_i(m)\big) > B.$$




شکل ۷: توابع $f$ و $g$ که در معادلات (۱۳) و (۱۴) تعریف شده اند.

از آنجایی که

$$\mathbb{E}[b\big(X_i(m)\big)] = \frac{B}{1+\epsilon}$$

در نتیجه

$$\mathbb{E}\big[\frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}}\frac{1}{n}\sum_{i=1}^{n}b\big(X_i(m)\big)\big] = \frac{B}{1+\epsilon}.$$

از قضیه قانون اعداد بزرگ میتوان نتیجه گرفت که با احتمال زیاد شرط هزینه برقرار خواهد شد.

## ۸ پیوست

در این پیوست نشان میدهیم که $C(B)$ بر حسب $B$ پیوسته است. در اینجا به قضیه زیر نیاز داریم:

**قضیه ۲۸** هر تابع محدب و یا مقعری که در بازه ای باز تعریف شده باشد، در این بازه باز پیوسته است. اگر این توابع در یک بازه بسته تعریف شده باشند، در تمام نقاط، بجز احتمالا در دو سر بازه، پیوسته هستند.

در اینجا اثبات را نمی آوریم، اما اثبات آن را میتوانید در اینترنت[۱۷] بیابید. بنابراین $C(B)$ بر حسب $B$ پیوسته است، بجز احتمالا در نقطه گوشه ای $B = B_{min}$. پیوستگی در نقطه $B = B_{min}$ را هم میتوان اثبات کرد، اما ما از آن اغماض میکنیم.

اما این سؤال مطرح میشود که آیا میتوان بصورت کلی در مورد پیوسته بودن توابعی که به شکل

$$F(r) = \max_{x:f(x)\leq r} g(x),$$

تعریف شده اند حرفی زد که در آن $f$ و $g$ توابعی پیوسته هستند. در ادامه مثال نقضی می آوریم که نشان میدهد که در صورتی که تابع $f$ خطی نباشد، لزوما $F(r)$ بر حسب $r$ پیوسته نیست.

[۱۷] http://planetmath.org/ContinuityOfConvexFunctions.html

**مثال ۲۹** [۱۸] فرض کنید که

$f, g : [0,1] \to \mathbb{R}$ بصورت زیر تعریف شده باشند:

$$f(x) = \begin{cases} 1 & 0 \le x \le 1/3 \\ 3-6x & 1/3 \le x \le 1/2 \\ 6x-3 & 1/2 \le x \le 2/3 \\ 1 & 2/3 \le x \le 1 \end{cases} \tag{۱۳}$$

و

$$g(x) = \begin{cases} 1-3x & 0 \le x \le 1/3 \\ 0 & 1/3 \le x \le 2/3 \\ 3x-2 & 2/3 \le x \le 1 \end{cases} \tag{۱۴}$$

این توابع در شکل ۷ نمایش داده شده اند. بسادگی میتوان دید که $F(r) = 0$ برای هر $r < 1$ ولی $F(1) = 1$ که نشان میدهد که $F$ در $r = 1$ پیوسته نیست.

[۱۸] این مثال متعلق به آقای پیام دلگشا است