نظریه اطلاعات کلاسیک_ بخش دوم

وحید کریمی پور_ دانشکده فیزیک _ دانشگاه صنعتی شریف

۲۷ فروردین ۱۴۰۴

۱ مقدمه

می خواهیم رشته هایی از 0 و 1 را از یک کانال مخابراتی که دارای نوفه است مخابره کنیم. ولی این کانال تحت تاثیر نوفه [۱] قرار دارد و هرعلامت 0 یا 1 کهارسال می کنیم بااحتمال q تبدیل به 1 یا 0 شده و با احتمال $1-q$ دست نخورده باقی می ماند. می خواهیم کاری کنیم که گیرنده پیام ها همچنان بتواند پیام های صحیح را ازپیام های دریافت شده استخراج کند. راهی که برای تعیین وتصحیح خطا باید پیش بگیریم آن است عنصر تکرار را به نوعی وارد پیام های خود کنیم. به عنوان مثال می توانیم بجای یک 0 سه تا 0 وبجای یک 1 سه تا 1 مخابره کنیم، یعنی از کد تکرار سه تایی استفاده کنیم

$$0 \longrightarrow 000 \qquad 1 \longrightarrow 111 \tag{۱}$$

واز گیرنده بخواهیم که با استفاده از قانون اکثریت تصمیم بگیرد که یک رشته سه تایی که دریافت کرده است درواقع چه رشته ای بوده است . درواقع احتمال خطا که قبلاً برابربود با q اینک کمترشده است، زیرااحتمال وقوع خطا اینک برابراست با احتمال وقوع دو برگردان و یا سه برگردان درکد سه تایی که اولی برابر است با $3q^2(1-q)$ ودومی برابراست با q^3. درنتیجه احتمال وقوع خطا برابرخواهد بود با

$$P_E = q^3 + 3q^2(1-q) \tag{۲}$$

[۱] Noise

که برای q های کوچک ازمرتبه q^2 است. البته کد تکرار تنها یکی از روش های مقابله با خطاست. مسئله اصلی این است که تعداد بیت های بیشتری ارسال کنیم تا این تعداد بیت های بیشتر (مثلا در کد هامینگ یا نظیر آن) برای تصحیح خطا مورد استفاده واقع شوند. این کار را اضافه کردن [۲] می گویند. یعنی اضافه کردن بیت هایی که حامل اطلاعات نیستند. البته بهایی برای این کاهش خطا پرداخت کرده ایم و آن این است که نرخ مخابره اطلاعات را پایین آورده ایم و از سه بیت برای مخابره یک بیت استفاده کرده ایم. دراین حالت نرخ مخابره اطلاعات یعنی R برابراست با $\frac{1}{3}$. اما این نرخ کم است. اما هنوز احتمال خطا وجود دارد اگر چه احتمال وقوع آن را نسبت به قبل کمتر کرده ایم؟ آیا می توانیم احتمال وقوع خطا را پایین تر بیاوریم؟ پاسخ اش مثبت است. بجای کد تکرار سه تایی می توانیم کد تکرار ۵ تایی یعنی کد زیر استفاده کنیم:

$$0 \longrightarrow 00000 \qquad 1 \longrightarrow 11111. \tag{۳}$$

فاصله این دو رشته از هم برابر با ۵ است و احتمال وقوع خطا برابر است با احتمال اینکه سه تا بیشتر از این بیت ها دچار خطا شوند . این خطا برابر است با

$$P_E = \binom{5}{3} q^3(1-q)^2 + \binom{5}{4} q^4(1-q) + q^5 \sim 10q^3 \tag{۴}$$

که به مراتب از میزان خطای قبلی کمتر است. اما برای کاهش خطا مجبور شده ایم که نرخ مخابره اطلاعات را بازهم پایین بیاوریم و آن را به $R = \frac{1}{5}$ برسانیم. آیا می توانیم کدی به کار ببریم که نرخ اش از این مقدار بالاتر باشد؟ برای این کار می توانیم به جای اینکه تک تک علائم 0 و 1 را کد کنیم بلوک های دوتایی از این علایم را کد کنیم. مثلا چیزی شبیه به این:

$$\begin{aligned} 00 &\longrightarrow 00000 \\ 01 &\longrightarrow 11100 \\ 10 &\longrightarrow 00111 \\ 11 &\longrightarrow 11011. \end{aligned} \tag{۵}$$

در این کد نیز فاصله هر دو رشته از یک دیگر برابر با سه است و بنابراین احتمال خطا از مرتبه q^3 است ولی نرخ مخابره پیام کمی بهتر شده است و از $\frac{1}{3}$ به $\frac{2}{5}$ رسیده است. به نظر می رسد که با کد کردن رشته های دوتایی به جای تک تک حروف الفبا توانسته ایم ضمن اینکه احتمال خطا را کاهش می دهیم نرخ را نیز کمی بالا ببریم. آیا می توانیم بهتر از این عمل کنیم؟ مثلا چطور است که بلوک های چهارتایی را کد کنیم، آنهم توسط کد هامینگ یعنی کد $[4, 7, 3]$. این کد فاصله ۳ دارد و چهار بیت را در ۷ بیت کد می کند. به این ترتیب نرخ مخابره را از $\frac{2}{5}$ نیز کمی بالاتر می

[۲] Adding Redundancy

برد و به $\frac{4}{7}$ می رساند، ضمن اینکه احتمال خطا را از همان مرتبه q^3 نگاه می دارد. ولی خب بازهم احتمال خطا وجود دارد و ما به هیچ وجه نمی خواهیم مخابره ای با احتمال خطا انجام دهیم چرا که از بین میلیون ها بیت مخابره شده ده ها هزار بیت کاملا به اشتباه مخابره خواهند شد. آنچه که ما می خواهیم این است که احتمال خطا را به سمت صفر میل دهیم. اما همانطور که در کد تکرار دیدیم به نظر می رسد که برای پایین آوردن احتمال خطا به صفر ما مجبوریم نرخ مخابره را نیز به صفر میل دهیم. آیا راهی برای برون رفت از این اشکال وجود دارد؟ پاسخ اش مثبت است. راه اش را در مثال های بالا دیدیم. راه اش این است که بلوک های بزرگ تر را کد کنیم. یعنی k بیت را به n بیت کد کنیم. در این صورت نرخ مخابره برابر خواهد بود با $R = \frac{k}{n}$ و ممکن است که با بزرگ تر گرفتن هر چه بیشتر k و n این نسبت به سمت یک مقدار غیر صفر میل کند و بسیار مهم تر از آن شاید بتوانیم همزمان احتمال وقوع خطا را نیز به صفر برسانیم. در این صورت حد $\lim_{n\longrightarrow\infty} \frac{k}{n}$ را ظرفیت کانال می نامیم. این در واقع محتوی قضیه دوم شانون یا قضیه کانال نوفه دار شانون [۳] است.

- **قضیه دوم شانون:** برای هر کانال کلاسیک یک ظرفیت مثل C وجود دارد که می بایست با توجه به مشخصات کانال آن را محاسبه کرد. در این صورت همواره می توان با نرخی کمتر از این مقدار ظرفیت مخابره اطلاعات انجام داد بدون اینکه مرتکب هیچ نوع خطایی بشویم. به عبارت دقیق تر به ازای هر دو عدد کوچک ϵ و δ همواره یک نوع کدگذاری خاص مثل C^* و یک عدد صحیح مثل $n*$ وجود دارد که اگر طول رشته ها را بزرگ تر از n^* بگیریم در این صورت به ازای آن کد گذاری خاص هر دو شرط زیر توامان برآورده می شوند:

$$R < C - \delta \qquad P_E(x) < \epsilon \qquad \forall x, \tag{۶}$$

که در آن $P_E(x)$ احتمال خطا در مخابره رشته x است.

نخست سعی می کنیم به صورت شهودی این قضیه را بفهمیم و سپس یک اثبات کامل ارائه خواهیم داد. فرستنده کانال را به پیروی از نامگذاری رایجی که در مکانیک کوانتومی وجود دارد آلیس می نامیم. گیرنده را نیز باب می نامیم. برای سادگی فرض کنید که منبع پیام هایی را از منبع X صادر می کند که الفبای آن حروف $\{0, 1\}$ هستند و احتمال وقوع 0 و 1 نیز یکسان است. به عبارت دیگر آنتروپی منبع دراین مثال برابراست با 1. هم چنین فرض کنید که کانال کلاسیک ما یک کانال ساده است با احتمالات شرطی زیر:

$$P(0|0) = 1 - q, \qquad P(1|1) = 1 - q. \tag{۷}$$

[۳] Shannon Noisy Channel Theorem

بنابراین هر بیت با احتمال q دچار خطا می شود، مستقل از اینکه مقدار آن صفر یا یک است. چنین کانالی را کانال متقارن دوتایی [۴] می گوییم. حال رشته های k بیتی را در نظر بگیرید که از این منبع وارد کانال می شوند. در این رشته ها ممکن است خطا رخ دهد. هر رشته در اثر خطا می تواند به 2^k رشته دیگر تبدیل شود. اما نکته این است که همه این رشته ها واقعا در اثر خطا با احتمال یکسان تولید نمی شوند. در اینجا بازهم به مفهوم اساسی رشته های نمونه می رسیم. به این معنا که اگر k رشته بزرگی باشد، به طور متوسط kq تا از بیت ها دچار خطا می شوند و هر رشته عموما به

$$\binom{k}{kq} \approx 2^{kH(q)}$$

تا رشته دیگر تبدیل می شود. در اینجا تابع $H(q)$ همان آنتروپی شانون برای خطا ست که برابراست با

$$H(q) = -q\log_2 q - (1-q)\log_2(1-q). \qquad (۸)$$

بنابراین اگر یک ناظر فرضی را تصور کنیم که بر کانال کلاسیک اشراف دارد و می بینید که یک رشته دچار خطا شده است، کافی است که شماره رشته تبدیل یافته را به باب اطلاع دهد تا او بتواند با اطلاعاتی که از رشته اولیه دارد، رشته دریافتی را تصحیح کند. برای این کار این ناظر فرضی می بایست $kH(q)$ بیت اضافه به باب ارسال کند. تا اینجا تعداد بیت هایی که برای مبادله بدون خطا حساب کردیم برابر شد با

$$k + kH(q).$$

اما ممکن است بیت های اضافه ای که ناظر فرضی برای گیرنده می فرستد خود نیز دچار خطا شوند. برای تصحیح آنها لازم است که $kH(q)^2$ بیت اضافه نیز به باب ارسال شود. در نتیجه برای اینکه امکان خطا به کلی از بین برود می بایست تعداد

$$n := k + kH(q) + kH^2(q) + kH^3(q) + \cdots = \frac{k}{1-H(q)}. \qquad (۹)$$

بیت به باب ارسال شود. بنابراین برای اینکه k بیت را بدون خطا بفرستیم مجبوریم $n = \frac{k}{1-H(q)}$ ارسال کنیم. نسبت این دو همان چیزی است که نرخ مبادله پیام نام دارد: در نتیجه نرخ مبادله پیام در چنین کانالی برابر است با:

$$C := \frac{k}{n} = 1 - H(q). \qquad (۱۰)$$

این بهترین نرخ مبادله اطلاعات است که آن را ظرفیت کانال می نامیم.

---

[۴] Binary Symmetric Channel

به یک روش دیگر نیز می توانیم این رابطه آخری را بفهمیم. آلیس n بیت به باب می فرستد. از این n بیت، $nH(q)$ بیت آن حامل اطلاعات نیست و فقط برای تصحیح خطا به کار رفته است، این بیت ها به باب می گویند که کدام بیت ها دارای خطا هستند و می بایست تصحیح شوند. باقیمانده بیت ها یعنی $n - nH(q)$ تای آنها برای کد کردن k بیت به کار رفته اند. بنابراین $k \leq n - nH(q)$ و در نتیجه

$$R_{max} := C = \frac{k}{n} \leq 1 - H(q). \qquad (۱۱)$$

باید تاکید کنیم که اثباتی که تا کنون ارائه کردیم، یک اثبات شهودی بود که هدف اش بیان کلی یک ایده بود. حال می توانیم اثبات دقیقی از آنچه که گفتیم ارائه کنیم. در این اثبات نیز فرض می کنیم که منبع $X = \{0, 1\}$ منبعی است با آنتروپی یک. در ادامه اثباتی را که در اینجا ارائه می دهیم به حالت کلی تعمیم خواهیم داد.

■ اثبات قضیه دوم شانون:

مجموعه تمام رشته های n بیتی دارای 2^n رشته است. نقاط این فضا نقاط یک شبکه ابرمکعبی n بعدی را پرکرده اند. اگر همه این نقاط را به عنوان کدرشته های خود به کاربریم به آسانی درطول عبوراز کانال یک کد رشته به یکی از کد رشته های همسایه اش یا کد رشته های نزدیکش تبدیل می شود و گیرنده واقعاً نمی فهمد که چه کد رشته ای برایش ارسال شده است. بنابراین راه مقابله با خطا آن است که سعی کنیم تنها تعدادی از این 2^n رشته را به عنوان کدرشته های واقعی خود یعنی کد رشته هایی که حامل اطلاعات هستند انتخاب کنیم. به عبارت دیگر کدرشته ها را بافاصله کافی از یکدیگر انتخاب کنیم تااحتمال وقوع خطا پایین بیاید. اگر مطابق شکل (۲) ، فقط 2^k رشته را انتخاب کنیم مثل این است که k بیت اطلاعات را در n بیت کد کرده ایم و ارسال می کنیم. برای چنین کدی نرخ مبادله اطلاعات برابر است با:

$$R := \frac{k}{n}. \qquad (۱۲)$$

روش کد کردن به این ترتیب است که حول هر کدام از 2^k تا رشته اصلی، یک کره به اندازه کافی بزرگ رسم می کنیم و هروقت نقطه ای درون این کره دریافت کنیم آن را به عنوان کد رشته ای که درمرکز آن قراردارد تعبیر وخطاهای بوجود آمده را تصحیح می کنیم. این کاربه طور طبیعی نرخ را پایین می آورد زیرازهمه نقاط شبکه استفاده نمی کنیم. حال می پرسیم که کد رشته ها را با چه فاصله ای انتخاب کنیم. دراین جا می بایست بین دو عامل متضاد موازنه ایجاد کنیم. اگر بخواهیم خطارا پایین بیاوریم می بایست فاصله کد رشته ها را از یکدیگر زیاد کنیم. ازطرفی این کارنرخ را پایین می آورد. حال هدف ما این است که ببینیم آیا می توانیم احتمال خطا را به سمت صفر میل دهیم و در عین حال نرخ بالا را محدود نگاه داریم. به عبارت دقیق تر هدف ما این است که بزرگترین نرخ ممکن را وقتی که خطا به سمت صفر میل می کند پیدا کنیم. آن بزرگترین نرخ ممکن همان ظرفیت است. خوب بیایید ببینیم به طور متوسط یک کد رشته به چند کدرشته نزدیک دیگر ممکن است تبدیل شود؟

در یک رشته n حرفی هر حرف با احتمال q ممکن است که برگردانده شود. بنابراین تعداد حرف های برگردانده شده به طور متوسط برابراست با nq. این تعداد حرف می تواند به $\approx 2^{nH(q)} \begin{pmatrix} n \\ nq \end{pmatrix}$ طریق در رشته n حرفی قرارگیرند. درنتیجه تعداد رشته های نزدیک به این رشته که ممکن است ازخطاهای بوجود آمده دراین رشته ایجاد شوند برابراست با $2^{nH(q)}$. دراین جا نیز ایده اصلی این است که اگر چه کل خطاهای ممکن در یک رشته n تایی برابراست با 2^n اما تعداد خطاهای متعارف یا نمونه از این مقدار کمتر است. در حقیقت اگر وقوع یک خطا را با بیت 1 و عدم وقوع آن را با بیت 0 نشان دهیم آنگاه تمام خطاهای ممکن تبدیل می شوند به رشته هایی از صفر و یک که تعدادشان برابراست با 2^n ولی نکته در این است که ما تنها می بایست به خطاهای نمونه یا $typical$ فکر کنیم که تعدادشان برابر است با $2^{nH(q)}$ که کمتر از مقدار کل است. تمامی قضایایی که در درس گذشته قاره تعداد و احتمال رشته های نمونه یاد گرفتیم در مورد خطاهای نمونه نیز صادق است.

درنتیجه یک نحوه کد گذاری خوب آن است که حول هررشته ناحیه ای درنظربگیریم که به طور متوسط تعداد رشته های داخل آن درحدود $2^{nH(q)}$ باشد. حال اگر نرخ مبادله برابربا $R = \frac{k}{n}$ باشد معنایش این است که تعداد کل کدرشته ها ها برابراست با $2^k = 2^{nR}$. چون تعدادکل رشته های ممکن برابراست با 2^n به این نتیجه می رسیم که می بایست شرط زیربرقرارباشد:

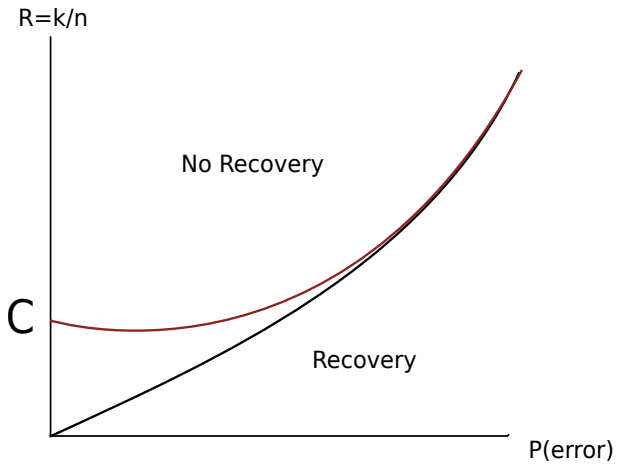
$$2^{nR}2^{nH} \leq 2^n, \tag{۱۳}$$

یا

$$R \leq 1 - H(q) =: C. \tag{۱۴}$$

کمیت $C := 1 - H(q)$ ظرفیت این کانال نامیده می شود واین رابطه بیان می کند که برای پرهیزازخطا نر خ مبادله اطلاعات حتما باید ازظرفیت کانال کمترباشد. در واقع آنچه که نشان داده ایم این است بر خلاف تصور اولیه می توان منحنی نرخ بر حسب خطا را از سقوط به نقطه صفر نجات داد و با توجه به این نکته که خطاهای نمونه خیلی کمتر از تعداد کل رشته های ممکن است می توانیم یک نوع کدکردن به کار ببریم که ضمن آنکه نرخ خطا را در حد مجانبی به صفر می رساند ولی نرخ مبادله اطلاعات را به سمت یک مقدار حدی میل می دهد. هر گونه مبادله اطلاعات با کمتر از این نرخ و بدون خطا ممکن است اما هرگونه تجاوز از این نرخ باعث می شود که هیچ گونه اطلاعاتی قابل بازیابی نباشد.

البته اثبات ما کامل نیست زیرا هنوز نشان نداده ایم که احتمال خطا واقعا به سمت صفر میل می کند. در اینجا می خواهیم نشان دهیم که حتما یک کد وجود دارد که با به کار بردن آن می توانیم نرخ مبادله خطا را هرچقدر که می خواهیم بهC نزدیک کرده و احتمال را هرچقدر که می



شکل ۱ : رابطه نرخ مبادله اطلاعات و نرخ خطا و مفهوم ظرفیت.

خواهیم کوچک کنیم. دقت کنید که در اینجا مسأله ما ساختن صریح این کد نیست بلکه می خواهیم نشان دهیم که یک چنین کدی وجود دارد. برای اثبات به ترتیب زیر پیش می رویم. فضای تمام های n بیتی را در نظر می گیریم. این فضا دارای 2^n نقطه است. حال به طور تصادفی 2^k نقطه را درون این فضا در نظر می گیریم. (درست مثل پرتاب کردن دارت به یک صفحه.) این حرف به این معناست که با احتمال یکنواخت این نقاط را انتخاب می کنیم. بنابراین نحوه کدکردن ما به همین سادگی است، یعنی به جای تعریف صریح کد (مثلا نوشتن کد خطی با یک ماتریس مولد یا ماتریس پاریته و نظایرآن) تنها 2^k نقطه را به طور تصادفی در این فضا اختیار می کنیم و این نقاط را به عنوان کد رشته های خود به کار می بریم. به این ترتیب می توانیم k بیت را در n بیت کد کنیم. این کد را یک کد تصادفی [۵] می نامیم و آن را با C_1 نشان می دهیم. دقت کنید که ممکن است بعضی از رشته ها کد به هم نزدیک انتخاب شوند که در این صورت احتمال رخ دادن خطا برای این کد رشته ها زیاد است. در عوض در بعضی موارد فاصله یک کد رشته می تواند از بقیه خیلی زیاد باشد که در این صورت احتمال رخ دادن خطا یعنی تبدیل این کد رشته به یک کد رشته دیگر خیلی کم است. هر بار که این نقاط را انتخاب می کنیم یک کد جدید به طور تصادفی مشخص می شود. این کد ها را با $\mathcal{C}_1$ ، $\mathcal{C}_2$ و نظایر آن نشان می دهیم. این آنزامبل از همه کدهای تصادفی در اثبات قضیه شانون نقش مرکزی ایفا می کند.

■ قضیه: به ازای هر مقدار کوچک ϵ و هر مقدار کوچک δ حتما یک کد $\mathcal{C}^*$ وجود دارد به قسمی که : احتمال خطا برای تمام رشته ها از ϵ کمتر باشد و نرخ مبادله پیام نیز هر چقدر که می خواهیم به ظرفیت نزدیک باشد به این معنا که شرط $C - R < \delta$ برقرار باشد.

برای اثبات این قضیه به ترتیب زیر عمل می کنیم. هر رشته ای که دریافت می کنیم به دورش یک کره به شعاع $(H(q) + \delta)$ رسم می کنیم.

[۵] Random Code

اصطلاحاً این به این معناست که مجموعه تمام نقاطی را در نظر می گیریم که دارای $2^{n(H(q)+\delta)}$ نقطه باشند. هرگاه یک کد رشته درون این کره قرار داشت می گوییم رشته ای که ارسال شده است همان نقطه بوده است و رشته دریافت شده را به همان رشته اصلاح می کنیم، شکل ۲. خطای بازگشایی وقتی رخ می دهد که رشته دیگری از جای دیگری درون کره مورد نظر ما بیفتد، شکل (؟؟). بجای اینکه به یک کد مشخص نگاه کنیم مجموعه تمام کدهای تصادفی را در نظر می گیریم و احتمال چنین خطایی را به طور متوسط حساب می کنیم. این متوسط روی تمام کدهای تصادفی و هم چنین روی تمام رشته ها در هر کدام از کدهاست. این احتمال خطا را با $\langle \overline{P} \rangle$ را نشان می دهیم. توجه به این نکته مهم است که این نماد در واقع نشان دهنده دو نوع متوسط پشت سرهم است که در بالا به آن اشاره شد. به طور دقیق تر برای یک کد مشخص مثل $\mathcal{C}$ داریم

$$\overline{P^{\mathcal{C}}} = \frac{1}{2^k} \sum_{x=1}^{2^k} P^{\mathcal{C}}(x) \qquad (۱۵)$$

که در آن $P^{\mathcal{C}}(x)$ احتمال این است که در کد $\mathcal{C}$ ما در بازگشایی رشته x دچار خطا شده باشیم. هم چنین علامت براکت برای محاسبه متوسط روی همه کدهای تصادفی اختیار شده است یعنی به ازای هر کمیت مثل O

$$\langle O \rangle = \frac{1}{\mathcal{Z}} \sum_{\mathcal{C}} O(\mathcal{C}), \qquad (۱۶)$$

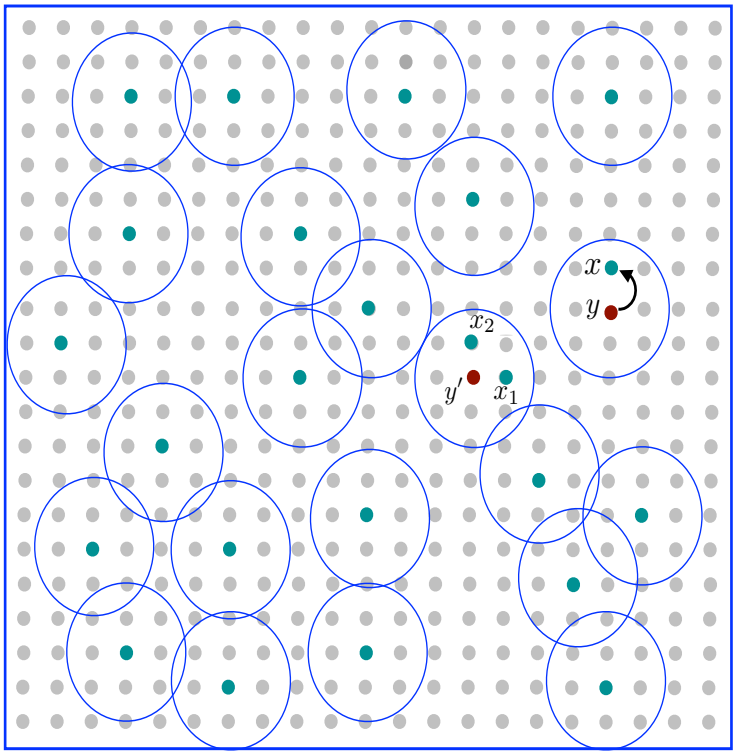
که در آن $\mathcal{Z}$ تعداد کدهای تصادفی است. حال می توانیم $\langle \overline{P} \rangle$ را حساب کنیم. این احتمال را بدون توجه به جزییات کدها می توان حساب کرد. می خواهیم ببینیم احتمال اینکه نقطه دیگری از یک کره دیگر به درون یک کره مورد نظر ما بیفتد چقدر است؟ در اثر خطا هر نقطه سبز رنگی می تواند شروع به حرکت کند و در جای دیگری از این فضا که شامل 2^n نقطه است بنشیند. اگر یک کره مشخص را حول نقطه ای مثل x در نظر بگیریم ممکن است نقطه ای از یک کره نزدیک حرکت کرده و درون این کره قرار گرفته باشد یا اینکه نقطه ای از یک کره دوردست دچار خطای زیاد شده و به درون این کره رسیده باشد. احتمال اینکه یک رشته (یک نقطه سبزرنگ) در اثر خطا (یا ولگشت) به درون این کره افتاده باشد برابر است با:

$$\frac{2^{n(H(q)+\delta)}}{2^n}. \qquad (۱۷)$$

تعداد کل رشته ها یا نقاطی که می توانسته اند دچار خطا شده باشند برابر است با $2^k = 2^{nR}$. بنابراین احتمال متوسط خطا برابر ا ست با:

$$\langle \overline{P} \rangle = 2^{nR} \times \frac{2^{n(H(q)+\delta)}}{2^n} = 2^{n(R-[1-H(q)]+\delta)} = 2^{n(R-C+\delta)} \qquad (۱۸)$$

آنچه که این استدلال مبتنی بر نسبت حجم ها را موجه می کند این است که در طرف چپ، دو نوع متوسط گرفته شده است یکی متوسط روی تمام رشته ها در هر کد و دوم متوسط روی تمام کدهای تصادفی. این رابطه نشان می دهد که اگر $R < C - \delta$ باشد، با انتخاب رشته های به



شکل ۲ : هر رشته کد شده با رنگ سبز نشان داده شده است. وقتی که باب یک کد رشته y (به رنگ قرمز) را دریافت می کند تمامی کد رشته هایی را که می توانسته اند به آن تبدیل شوند در نظر می گیرد. این کد رشته ها کره ای به شعاع $nH(q)$ را تشکیل می دهند. اگر درون این کره یک کد رشته معتبر (به رنگ سبز) وجود داشته باشد، آن کد رشته را به عنوان کد رشته معتبر فرستاده شده توسط آلیس تلقی خواهد کرد. خطا وقتی رخ می دهد که بیش از یک کد رشته درون این کره قرار بگیرد. احتمال اینکه هیچ کد رشته ای درون این کره قرار نگیرد ناچیز است.

اندازه کافی بزرگ یعنی n های بزرگ می توانیم احتمال متوسط خطا را روی همه کدهای تصادفی به سمت صفر میل دهیم. به عبارت دیگر به

ازای هر ϵ و هر δ خواسته شده کافی است که قرار دهیم:

$$2^{n(R-C+\delta)} \leq \epsilon \qquad (۱۹)$$

و از آنجا

$$n(R-C+\delta) \leq \log \epsilon \qquad (۲۰)$$

و در نتیجه

$$n \geq \frac{|\log \epsilon|}{C-\delta-R}. \qquad (۲۱)$$

این رابطه می گوید که هر چه بخواهیم ϵ را کوچکتر و یا R را به C نزدیک تر کنیم می بایست از رشته های طولانی تری استفاده کنیم.

■ تمرین:اگر بخواهیم که متوسط احتمال خطا کمتر از 0.001 شود و$R \leq C - 0.01$باشد طول کد رشته ها حداقل چقدر باید باشد؟ اگر بخواهیم با همین نرخ اطلاعات را مبادله کنیم ولی متوسط احتمال خطا را به زیر 0.00001 برسانیم حداقل طول کد رشته ها چقدر خواهد بود؟

استدلال بالا نشان می دهد که احتمال خطا را به طور متوسط می توان از هر مقداری که بخواهیم کمتر کنیم. ولی این خطا در واقع یک خطای متوسط دوگانه است به این معنا که یک بار روی تمامی کد رشته های درون یک کد متوسط گرفته ایم و باردیگر روی تمام کدهای تصادفی $C_1, C_2, \cdots$ متوسط گرفته ایم. به طور دقیق تر ثابت کرده ایم که می توان با انتخاب طول مناسب برای کد رشته ها نرخ مبادله اطلاعات را هرچقدر که می خواهیم به ظرفیت کانال نزدیک کنیم آنچنان که متوسط احتمال خطا از هر مقداری که می خواهیم کمتر باشد یعنی

$$\langle \overline{P} \rangle \leq \epsilon. \qquad (۲۲)$$

اما از رابطه ۲۲ نتیجه می شود که حتما یک کد تصادفی ای مثل C وجود دارد که برای آن رابطه زیر برقرار است:

$$\overline{P^{\mathcal{C}}} \leq \epsilon, \qquad (۲۳)$$

چرا که اگر همه کدهای تصادفی احتمال خطایشان از ϵ بیشتر بود، هیچگاه متوسط خطا روی همه کدهای تصادفی از ϵ کمتر نمی شد. اما این عبارت متوسط خطا را روی تمام کدرشته ها نشان می دهد. ممکن است برای بعضی از کد رشته ها خطا بیشتر از ϵ و برای بعضی دیگر کمتر از ϵ باشد. این کد البته مطلوب ما نیست چرا که ما کدی می خواهیم که احتمال خطا برای تمام کدرشته هایش از ϵ کمتر باشد. برای رسیدن به این

کد تنها کافی است که کد را کمی خلوت کنیم و بعضی از کد رشته ها را حذف کنیم. تنها یک نگرانی وجود دارد و آن اینکه با خلوت کردن کد و دور ریختن این نوع رشته ها تعداد خیلی کمی کد رشته در کد باقی بماند و نرخ به شدت پایین بیاید یا حتی به صفر برسد. نشان می دهیم که چنین نیست. در واقع نشان می دهیم که تنها با دور ریختن کمتر از نیمی از کد رشته ها (یا رشته ها) می توانیم به کدی برسیم که خطا برای تک تک کد_رشته ها از 2ϵ کمتر باشد و چون 2ϵ نیز یک مقدار بی نهایت کوچک است به هدف خود دست پیدا می کنیم. برای این کار بازهم از لم اول چبیشف استفاده می کنیم.

احتمال خطای مربوط به کد رشته اول را با $P^{\mathcal{C}}(x_1)$ ، خطای مربوط به کد رشته دوم را با $P^{\mathcal{C}}(x_2)$ نشان می دهیم و همینطور تا آخر. در این صورت می دانیم که متوسط این خطا ها از ϵ کمتر است. یعنی

$$\overline{P^{\mathcal{C}}} = \frac{1}{2^{nR}} \sum_{x=1}^{2^{nR}} P^{\mathcal{C}}(x) \leq \epsilon. \qquad (۲۴)$$

اگر x را به عنوان یک متغیر تصادفی ببینیم و $P^{\mathcal{C}}(x)$ را به عنوان تابعی از متغیرتصادفی x در نظر بگیریم، می توانیم لم چبیشف را برای آن به کار ببریم و بنویسیم:

$$Prob(P^{\mathcal{C}}(x) \geq 2\epsilon) \leq \frac{\overline{P^{\mathcal{C}}}}{2\epsilon} \leq \frac{\epsilon}{2\epsilon} = \frac{1}{2}. \qquad (۲۵)$$

معنای این حرف این است که کمتر از نصف تعداد کل کد رشته ها احتمال خطایشان از 2ϵ بیشتر است و اگر این تعداد از کد رشته ها را دور بریزیم بقیه همه احتمال خطایشان از 2ϵ کمتر است و این همان چیزی است که می خواستیم ثابت کنیم. بنابراین با دور ریختن کمتر از نیمی از کد رشته ها به کدی مثل C^* می رسیم که احتمال خطا برای تمام کد رشته هایش از 2ϵ کمتر است. کد جدیدی که ساخته ایم تعداد کد رشته هایش برابر است با:

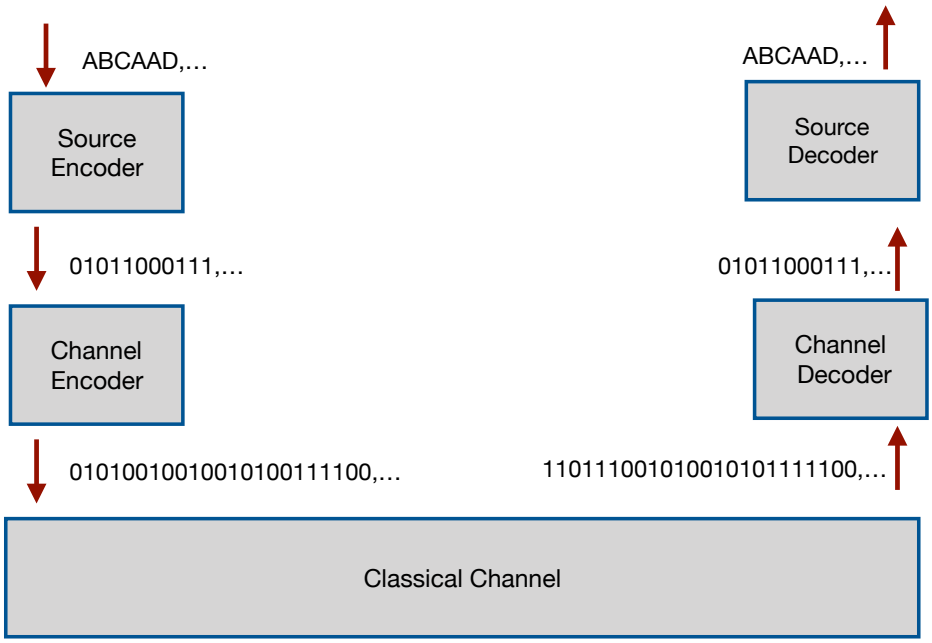
$$\frac{2^{nR}}{2} = 2^{n(R-\frac{1}{n})} \qquad (۲۶)$$

و این به این معناست که نرخ جدید در حد n های بزرگ با نرخ قبلی برابر است. بنابراین ثابت کردیم که یک کد وجود دارد که نرخ مبادله اطلاعات را به حد شانون یعنی ظرفیت کانال می رساند و در عین حال خطا را هم هر چقدر که بخواهیم کم می کند.

- تمرین: این تمرین را می توان جایگزین استفاده از قضیه چبیشف در استدلال بالا کنید. در واقع نمونه ساده ای از قضیه چبیشف است. برای اثبات آن از برهان خلف استفاده کنید. کلاسی را در نظر بگیرید که دارای یکصد دانشجوست. معدل کلاس در یک درس بخصوص برابر است با۱۰. نشان دهید که تعداد دانشجویانی که نمره بیش از ۲۰ گرفته اند نمی تواند از ۵۰ بیشتر باشد.

۲ تعریف ظرفیت درحالت کلی

ما تا کنون یک منبع ساده با آنتروپی یک در نظر گرفتیم. برای یک بحث دقیق تر، بهتر است یکبار دیگر به کل یک کانال مخابراتی کلاسیک از ابتدا تا انتها نگاه کنیم، شکل (۳).

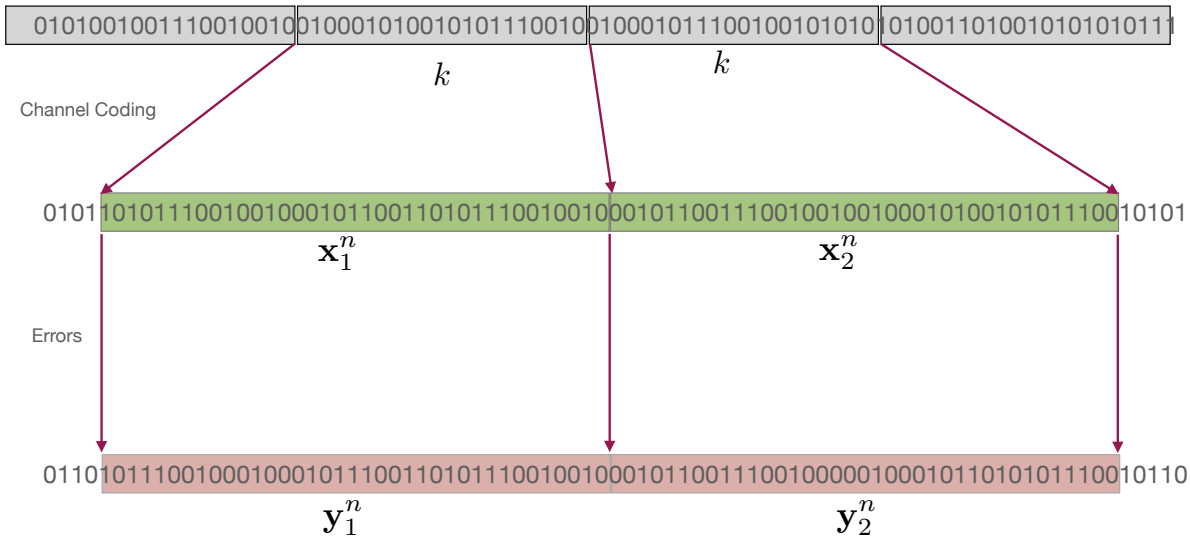


شکل ۳: نمای کلی یک کانال کلاسیک. اجزای این کانال و نقش هر کدام در متن توضیح داده شده است.

پیام اولیه نخست وارد یک کدکننده منبع [۶] می شود. کار این کدکننده این است که متن را مطابق با آنچه که در درس قبل دیدیم، فشرده کند. در بهترین حالت این فشردگی قرار است به حد شانون برسد اگرچه ممکن است در عمل این کار انجام نشود. دقت کنید که این کد کننده الزاما حروف را به تعداد بیت های یکسان کد نمی کند. از آنجا که هر چه طول کلمات یا رشته ها بلندتر باشد به حد شانون نزدیک می شویم، پس کد کننده منبع هم تک تک حرف ها را کد نمی کند بلکه رشته های بلند را کد می کند یا به اصطلاح فشرده می کند. این رشته ها سپس وارد کد

[۶] Source Encoder

کننده کانال می شوند که به آنها افزودنی [۷] اضافه می کند تا با بتواند خطاهای احتمالی ایجاد شده را آشکار کرده و اصلاح کند. بنابراین بلوک های k تایی به بلوک های به طول n کد می شوند، شکل (۴).



شکل ۴: هر رشته از حروف به طول k به یک رشته از حروف به طول n کد می شود. طول رشته ها هرچه بیشتر باشد بیشتر به ظرفیت کانال و احتمال خطای کم نزدیک می شویم.

مجموعه تمام این کد رشته های به طول n را با $X^{(n)}$ نمایش می دهیم. داریم

$$X^n = \{\mathbf{x}_1^n, \mathbf{x}_2^n, \cdots \mathbf{x}_\Omega^n\} \tag{۲۷}$$

که در آن Ω مجموعه تمام این کدرشته های به طول n است. از آنچه که قبلا درمورد رشته های نمونه خوانده ایم می دانیم که

$$|\Omega| = 2^{HX^{(n)}} \leq 2^n. \tag{۲۸}$$

هدف آلیس و باب این است که با بزرگ کردن هر چه بیشتر n احتمال خطا را هر چقدر که می توانند به سمت صفر میل دهند، ضمن آنکه $R := \frac{k}{n}$

[۷] Redundancy

را هرچقدر که می توانند زیاد کنند. کانال کلاسیک نیز با مجموعه احتمالات شرطی

$$P(\mathbf{y}^n|\mathbf{x}^n \tag{۲۹}$$

تعریف می شود. حال تعداد 2^k رشته بلند n تایی از نوع

$$\mathbf{x}^n$$

را در نظر بگیرید که قرار است به باب مخابره شوند و باب می بایست خطای احتمالی آنها را آشکار و تصحیح کند. این رشته ها در عبور از کانال به رشته های

$$\mathbf{y}^n$$

به طول n تبدیل می شوند. تعداد این رشته ها البته دیگر $2^{H(X^n)}$ نیست چون یک رشته ممکن است در اثر خطا به رشته های متعدد دیگر تبدیل شود. در واقع تعداد ا کلیه رشته هایی که در اثر خطا ممکن است به دست باب برسند برابر است با $2^{H(Y^n)}$، ولی نکته مهم آن است که تعداد رشته های معتبر (رشته های اولیه) هم چنان $2^{H(X^n)}$ است. حال سوال این است که باب چگونه رشته ها را بازگشایی می کند. درست مثل حالت ساده قبل، به دور یک کد رشته دریافتی مثل $\mathbf{y}^n \in Y$ مجموعه ای (کره مانند) در نظر می گیریم که شامل همه رشته های نمونه ای باشد که یک کد رشته در اثر خطا می توانسته به $\mathbf{y}^n$ تبدیل شود، شکل (۲). این تعداد را با V نمایش دهیم. هرگاه یک کد رشته از X درون این کره باشد، کد رشته y را به همان کد رشته بازگشایی می کنیم (می گوییم کد رشته فرستاده شده x بوده است که به y تبدیل شده است.) احتمال اینکه هیچ کدرشته ای درون این کره نباشد بسیار ناچیز است. خطا وقتی رخ می دهد که بیش از یک کد رشته درون این کروه قرار بگیرد. بنابراین برای تصحیح خطا انتظار داریم که کره ها با یک دیگر تداخل نکنند و هر کدام نیز بیش از یک کد رشته نداشته باشند. به عبارت دیگر می بایست داشته باشیم:

$$2^k \times V \leq 2^{H(X^n)}. \tag{۳۰}$$

اما نشان می دهیم V چیزی نیست جز $2^{H(X^n|Y^n)}$. بنابراین بدست می آوریم:

$$2^k \times 2^{H(X^n|Y^n)} \leq 2^{H(X^n)} \tag{۳۱}$$

و در نتیجه

$$2^k \leq 2^{H(X^n)-H(X^n|Y^n)} = 2^{H(X^n)+H(Y^n)-H(X^n,Y^n)}. \tag{۳۲}$$

با توجه به تعریف نرخ کانال یعنی $R := \frac{k}{n}$ خواهیم داشت:

$$R \leq \frac{H(X^n) + H(Y^n) - H(X^n, Y^n)}{n} = \frac{I(X^n : Y^n)}{n}. \quad (۳۳)$$

بنابراین در نگاه اول به نظر می رسد که ظرفیت یک کانال کلاسیک در حالت کلی عبارت است از:

$$C := \lim_{n \longrightarrow \infty} \frac{I(X^n : Y^n)}{n}. \quad (۳۴)$$

اما هنوز دو قدم کوچک دیگر باقی مانده است. نخست آنکه با فرض اینکه حرف های درون رشته های به طول n از یک دیگر مستقل اند و کانال نیز به طور مستقل روی آنها عمل می کند می توانیم بنویسیم

$$I(X^n : Y^n) = nI(X, Y)$$

و در نتیجه محاسبه ظرفیت تنها شامل محاسبه ای در مورد تک تک حرف ها خواهد بود، یعنی

$$C := I(X : Y \quad (۳۵)$$

و نکته دوم اینکه این کمیت به شکل فعلی بستگی به منبع ورودی یعنی X دارد و کمیتی کاملا وابسته به کانال نیست. بنابراین برای محاسبه ظرفیت می بایست مقدار بیشینه آن روی تمام منبع های ورودی حساب کرد. به این ترتیب عبارت نهایی برای ظرفیت کانال کلاسیک به شکل زیر بدست می آید.

$$\boxed{C := Max_X I(X : Y)}$$

باقی استدلال دقیق کاملا همانند حالت ساده ای است که با منبع ساده با الفبای صفر و یک بیان کردیم و هیچ چیز متفاوتی با آن حالت ندارد. به همان ترتیب قبل و دقیقا با همان تعریف ها می توان مقادیر بی نهایت کوچک ϵ و δ را به استدلال اضافه کرد، نخست آنزامبلی از کدهای تصادفی را در نظر گرفت و سپس نشان داد که حتما یک کد خاص وجود دارد که قضیه شانون برای برقرار می شود. باقی می ماند که V را حساب کنیم. نخست توجه می کنیم که رشته ای که در اثر خطا تولید می شود خود یک رشته نمونه به طول n است. این رشته دارای $nP(y_1)$ تا حرف y_1 است، $nP(y_2)$ تا حرف y_2 ، $nP(y_3)$ تا حرف y_3 است و همین طور تا آخر. حال توجه خود را به یکی از این حرف ها مثل y_j معطوف می کنیم. تعداد آنها $nP(y_j)$ تاست. از خود می پرسیم که چند تا از این حرف ها می توانسته است به اشتباه تولید شده باشد؟ پاسخ اش این است. هر رشته ای دارای $nP(x_i)$ تا x_i است و هر کدام از این x_i ها می توانسته با احتمال $P(y_j|x_i)$ به y_j تبدیل شده باشد. تعداد راه هایی که این اتفاق می توانسته بیفتد برابر است با:

$$\frac{[nP(y_j)]!}{\prod_{i=1}^{M}[nP(y_j|x_i)p(x_i)]!} \tag{۳۶}$$

و یا

$$\frac{[nP(y_j)]!}{\prod_{i=1}^{M}[nP(y_j,x_i)]!}. \tag{۳۷}$$

اگر همه y_j ها را در نظر بگیریم، تعداد رشته های نمونه ای که یک رشته از نوع x^M می توانسته به آن تبدیل شود برابر است با:

$$V=\prod_{j=1}^{M'}\frac{[nP(y_j)]!}{\prod_{i=1}^{M}[nP(y_j,x_i)]!}. \tag{۳۸}$$

برای رشته های بلند از تقریب استرلینگ استفاده می کنیم و پس از ساده کردن جملات درمی یابیم که:

$$\log V=n\sum_{j=1}^{M'}\left[P(y_j)\log P(y_j)-\sum_{x_i=1}^{M}P(y_j,x_i)\log P(y_j,x_i)\right] \tag{۳۹}$$

و یا

$$\log V=n\left[H(X,Y)-H(Y)\right], \tag{۴۰}$$

و این همان چیزی بود که می خواستیم ثابت کنیم.

۳ اصلاحیه

$$C:=Max_{P(x)}I(X;Y) \tag{۴۱}$$

$$I(X;Y)=H(X)+H(Y)-H(X,Y)=H(Y)-H(Y|X) \tag{۴۲}$$

۱.۳ کانال متقارن دوتایی

$$H(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_x P(x)H(Y|x) = H(Y) - \sum_x P(x)H(p)$$
$$H(Y) - H(p) \leq 1 - H(p) \quad (۴۳)$$

$$C = 1 - H(p) \quad (۴۴)$$

۲.۳ کانال پاک کننده دوتایی

$$C = Max_{P(x)} H(X;Y) = Max_{P(x)} \Big[H(Y) - H(Y|X) \Big] = Max_{P(x)} H(Y) - H(\alpha) \quad (۴۵)$$

$$H(Y) = H(Y,E) = H(E) + H(Y|E) \quad (۴۶)$$

$$p_1 = (1-\pi)(1-\alpha), \quad p_2 = \alpha, \quad p_3 = \pi(1-\alpha) \quad (۴۷)$$

$$H(Y) = H(p_1, p_2, p_3) = H(\alpha) + (1-\alpha)H(\pi) \quad (۴۸)$$

$$\begin{aligned} C &= Max_{P(x)} H(Y) - H(\alpha) \\ &= Max_{\pi} (1-\alpha)H(\pi) + H(\alpha) - H(\alpha) \\ &= 1 - \alpha \end{aligned} \quad (۴۹)$$

۴ خواص ظرفیت یک کانال

■

$$C \geq 0$$
$$C \leq \log(|X|),$$
$$C \leq \log(|Y|). \qquad (۵۰)$$

■ $I(X:Y)$ یک تابع پیوسته و محدب از $P(x)$ است.

۵ مسئله ها:

■ نشان دهید که برای سه متغیر تصادفی $A,\ B,\ C$ رابطه زیر برقرار است:

$$H(A,B,C) + H(B) \leq H(A,B) + H(B,C) \qquad (۵۱)$$

■ کارکردن با آنتروپی شانون به دلیل وجود لگاریتم معمولا سخت است. می توان بجای آن از کمیت های جایگزین زیر استفاده کرد:

$$\begin{aligned} L &= \sum_i p_i(1-p_i) \\ T &= \frac{1-\sum_i p_i^q}{1-q} \\ R &= \frac{\log(\sum_i p_i^q)}{1-q}. \end{aligned} \qquad (۵۲)$$

این ها به ترتیب آنتروپی خطی، آنتروپی تسالیس [۸] و آنتروپی رنی [۹] خوانده می شوند. نشان دهید که:

الف: هر کدام از این آنتروپی ها مثبت است و مقدارش تنها برای وقتی که یکی از احتمالات غیرصفر باشد، برابر با صفر می شود.

ب: هرکدام از این آنتروپی ها مقدار بیشینه خود را برای تابع توزیع یکنواخت اختیار می کند.

پ: نشان دهید که

$$H_e \geq L \qquad (۵۳)$$

[۸] Tsaliss Entropy
[۹] Renyi Entropy

و تساوی تنها وقتی برقرار می شود که $H_e = L$ باشد. در اینجا منظور از H_e عبارت زیر است:

$$H_e = -\sum_i p_i \log p_i \qquad (۵۴)$$

- فرض کنید که N توپ یکسان بین دو جعبه A و B بنا بر یک تابع توزیع معین به اشتراک گذاشته اند. به ازای کدام تابع توزیع، تابع $H(A:B)$ بیشترین مقدار خود را اختیار خواهد کرد. اگر توپ ها تمیز پذیر باشند، مقدار بیشینه $H(A:B)$ چه خواهد بود؟

- یک کانال کلاسیک سه حرف ورودی (a_1, a_2, a_3) و سه حرف خروجی (b_1, b_2, b_3) دارد. کانال با احتمالات شرطی زیر تعریف می شود:

$$\begin{aligned} &P(b_1|a_1) = 1 \\ &P(b_2|a_2) = 1 - q = P(b_3|a_3) \\ &P(b_2|a_3) = q = P(b_3|a_2). \end{aligned} \qquad (۵۵)$$

الف: ظرفیت کانال را حساب کنید.

ب: برای وقتی که پارامتر q یکی از مقادیر $0,\ 1/2\ ,1$ است، نشان دهید که چگونه می توان نرخ مبادله اطلاعات را به ظرفیت کانال رساند.

- همیشه لازم نیست که تعداد حروف ورودی و خروجی یک کانال کلاسیک با هم برابر باشند. یک کانال کلاسیک در نظر بگیرید که حروف ورودی آن (a_1, a_2) و حروف خروجی آن $(b_1,\ b_2,\ b_3,\ b_4)$ باشند. کانال نیز با احتمالات شرطی زیر تعریف می شود:

$$\begin{aligned} P(b_1|a_1) = P(b_2|a_1) &= \frac{1}{3}, \\ P(b_3|a_2) = P(b_4|a_2) &= \frac{1}{3}, \\ P(b_3|a_1) = P(b_4|a_1) &= \frac{1}{6}, \\ P(b_1|a_2) = P(b_2|a_2) &= \frac{1}{6}. \end{aligned} \qquad (۵۶)$$

ظرفیت این کانال را حساب کنید.

- یک پیام از بلوک های به طول هفت بیت صفر و یک تشکیل شده است. یک کانال به شکل زیر در این پیام اختلال ایجاد می کند: یا هیچ نوع خطایی رخ نمی دهد که احتمال آن $\frac{1}{8}$ است یا با احتمال $\frac{1}{8}$ در یکی از بیت های بلوک خطا ایجاد می کند.

الف: ظرفیت کانال را حساب کنید.

ب: آیا می توان شیوه ای را پیشنهاد کنید که در آن نرخ مبادله اطلاعات با ظرفیت کانال مساوی شود؟

■ در یک پیام که می دانیم توسط کد همینگ ایجاد شده است کلمات زیر را دریافت کرده ایم:

$$\begin{aligned} &(a) \quad 1110000 \\ &(b) \quad 1000101 \\ &(c) \quad 0100110 \\ &(d) \quad 0110010. \end{aligned} \tag{۵۷}$$

در هر مورد مشخص کنید که کلمه ارسال شده چه کلمه ای بوده است؟