نظریه اطلاعات کلاسیک ـ بخش دوم

وحیدکریمی پور ـ دانشکده فیزیک ـ دانشگاه صنعتی شریف

۲۴ اسفند ۱۳۹۳

۱ مقدمه

درفصل های قبل دیدیم که می توان پیام های یک منبع را که از کلمات $X = \{x_1, x_2, \cdots x_{2^n}\}$ تشکیل شده است را می توان باکدگذاری مناسب یعنی کدگذاری بلوکی چنان مخابره کرد که بجای n بیت به ازای هرکلمه، $nH(X) \leq n$ بیت مخابره کنیم. این نتایج البته درحد n های بزرگ دقیق هستند. تمام این حرف ها درصورتی بود که ازنوفه داخل کانال صرف نظرکنیم و فرض کنیم که پیام بدون هیچ نوع تغییری به مقصد می رسد. مطالب بالا محتوی قضیه شانون را درباره کدگذاری بدون نوفه تشکیل می دهند. دراین درس می خواهیم درباره کد گذاری درحضورنوفه بحث کنیم و نهایتا قضیه شانون درباره کد گذاری درحضورنوفه را بیان کنیم. دراین درس هم چنین بامفهوم مهم ظرفیت کانال آشنا خواهیم شد ودرحالت های خیلی ساده ای مقدارآن را محاسبه می کنیم.

۲ موازنه بین پایین آوردن خطا وبالابردن نرخ انتقال اطلاعات

ازاین به بعد فرض می کنیم که کلمات منبع خود را به طریق بهینه ای کدگذاری کرده ایم وتنها می خواهیم کد کلمه هارا که رشته هایی از 0 و 1 هستند مخابره کنیم. فرض می کنیم که کانالی که پیام های خود را از طریق آن ارسال می کنیم تحت تاثیرنوفه قرار می گیرد 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.

۱.۲ یک تذکرمهم

آنچه که درمثال بالاگفته شد یک نوع کد موسوم به کد تکرار سه تایی برای تشخیص وتصحیح خطاست و باکدگذاری کلمات فرق می کند. ازاین به بعد دراین درس هروقت که ماازکلمه کدگذاری استفاده می کنیم منظورمان این نوع کدگذاری است. به عبارت دیگر فرض می کنیم که کلمات منبع X که می توانند حروف یک زبان مثل فارسی یاانگلیسی یاعلائم ریاضی باشند به نحو بهینه ای تبدیل به رشته های 0 و 1 شده اند و هدف ما مخابره این رشته ها ازدرون کانال است. درانتهای کانال رشته های دریافت شده نخست برای تعیین وتصحیح خطاها بررسی شده وپس ازاصلاح به کلمات زبان موردنظربازگشایی می شوند.

راهی که برای تعیین وتصحیح خطاباید پیش گرفت آن است عنصر تکرار را به نوعی وارد پیام های خود کنیم. به عنوان مثال می توانیم بجای مخابرهیک 0 سه تا 0 وبجای یک 1 سه تا 1 مخابره کنیم 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 (جدول 1.2).
درواقع احتمال خطا که قبلاً برابربود با q اینک کمترشده است، زیرااحتمال وقوع خطا اینک برابراست بااحتمال 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 $q^3+3q^2(1-q)$ که برای q های کوچک ازمرتبه q^2 است. 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 $\frac{1}{3}$. درحالت کلی اگر از n بیت برای مخابره 2^k پیام استفاده کنیم (که درنبود نوفه می توانست

برای مخابره 2^n پیام مورد استفاده قرارگیرد) می گوییم که نرخ مخابره اطلاعات برابراست

$$R := \frac{k}{n}. \qquad (۱)$$

مسلم است که‌دراین نوع کدگذاری می توان بازهم احتمال خطا را کاهش داد و درحد ایده آل آن را به صفر نزدیک کرد ولی دراین صورت نرخ R نیزبه سمت صفرمیل خواهد کرد.

- تمرین: فرض کنید که کد تکرار n تایی را به کار می بریم که در آن n یک عدد فرد است. نرخ مخابره اطلاعات برابر است با $R = \frac{1}{n}$. در این کد احتمال خطا را بر حسب n حساب کنید. سپس منحنی نرخ برحسب احتمال خطا را رسم کنید.

- تمرین: در کد هامینگ که با مشخصات $[7, 4, 3]$ تعیین می شود نرخ مخابره اطلاعات و هم چنین احتمال خطا را تعیین کنید.

آنچه که مهم است این است که ما نمی خواهیم اطلاعات را با احتمال خطای کم مخابره یا در جایی ذخیره کنیم. هدف ما آن است که این کار را با نرخ خطای صفر انجام دهیم یا به عبارت دقیق تر مطلوب ما آن است که این نرخ خطا را هر چقدر که می خواهیم بتوانیم به صفر نزدیک کنیم. حال سوالی که باآن مواجه هستیم آن است که آیا اصولاً می توان ازیک کانال نوفه دار با نرخ محدود و غیرصفر اطلاعاتی را عبورداد به نحوی که احتمال وقوع خطا لااقل درحد مجانبی (برای رشته های بزرگ) به سمت صفر میل کند؟ آیا یک نوع کدگذاری وجود دارد که این امکان را برای ما فراهم آورد؟ آیا نرخ بیشینه ای وجود دارد که هرنوع مخابره اطلاعات با بیش از آن نرخ بدون خطا غیرممکن باشد؟ این ها سوالاتی است که درقضیه دوم شانون پاسخ داده شده است. قبل از بیان قضیه کلی سعی می کنیم که دو مثال ساده را بررسی کنیم.

نخست حالت ساده ای را درنظرمی گیریم که منبع X پیام هایی را ارسال می کند که درآنهااحتمال وقوع 0 و 1 یکسان باشد. به عبارت دیگر آنتروپی منبع دراین مثال برابراست با 1. فرض کنید که پیام خود را دررشته های n بیتی ارسال کنیم . 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 کدکلمه های خود به کاربریم به آسانی درطول عبوراز کانال یک کد کلمه به یکی از کد کلمه های همسایه اش یا کد کلمه های نزدیکش تبدیل می شود و گیرنده واقعاً نمی فهمد که چه کد کلمه ای برایش ارسال شده است. بنابراین راه مقابله با خطا آن است که سعی کنیم که کدکلمه هارا بافاصله کافی از یکدیگر انتخاب کنیم تااحتمال وقوع خطا پایین بیاید.

∘ ∘ ∘	0
۱ ۱ ۱	1

جدول ۱: کد تکرار سه تایی

درواقع وقتی که یک کدکلمه راارسال می کنیم حول آن یک کره به اندازه کافی بزرگ رسم می کنیم و هروقت نقطه ای درون این کره دریافت کنیم آن را به عنوان کد کلمه 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 کلمه هارا 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 کلمه هاراازیکدیگر زیادکنیم. ازطرفی این کارنرخ را پایین می آورد. خوب بیایید ببینیم به طور متوسط یک کد کلمه به چند کلمه نزدیک دیگر ممکن است تبدیل شود؟

در یک کلمه 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 $\binom{n}{nq} \approx 2^{nH(q)}$ طریق در رشته n حرفی قرارگیرند. درنتیجه تعداد رشته های نزدیک به این کلمه که ممکن است ازخطاهای بوجود آمده دراین کلمه 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 خیلی کمتر از مقدار کل است. در اینجا تابع $H(q)$ همان آنتروپی شانون برای خطا ست که برابراست با

$$H(q) = -q \log_2 q - (1-q) \log_2 (1-q). \qquad (۲)$$

تمامی قضایایی که در درس گذشته در باره تعداد و احتمال رشته های نمونه یادگرفتیم در مورد خطاهای نمونه نیز صادق است.

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

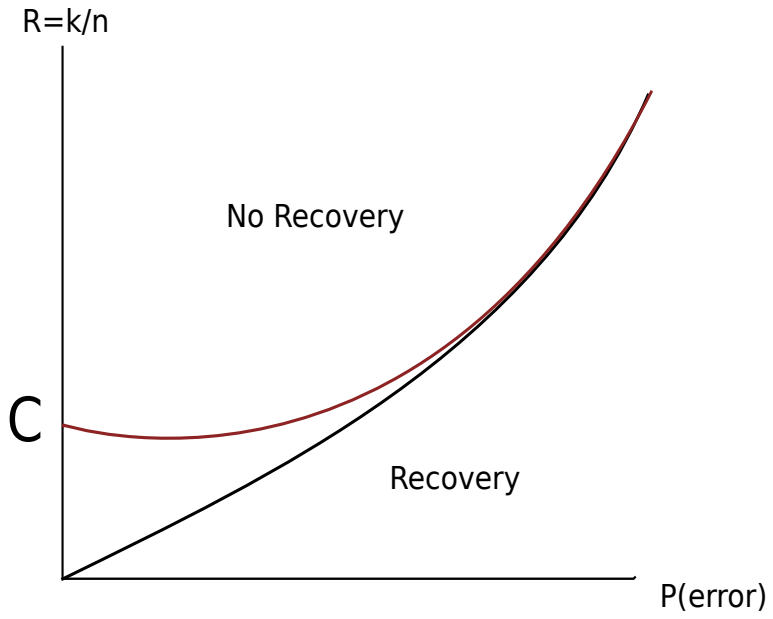
$$2^{nR}2^{nH} \leq 2^n, \tag{۳}$$

یا

$$R \leq 1 - H(q) =: C. \tag{۴}$$

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.

حال بایک سوال اساسی مواجه می شویم و آن اینکه آیا واقعا نوعی کدگذاری وجود دارد که بتوان بااستفاده از آن نرخ مبادله اطلاعات رادرحد مجانبی $n \longrightarrow \infty$ به حداکثر ممکن یعنی به C رساند؟ پاسخ این سوال مثبت است . در واقع نشان می دهیم که حتما یک کد وجود دارد که می تواند نرخ مبادله بدون خطای اطلاعات را هرچقدر که می خواهیم به ظرفیت C نزدیک کند. به عبارت دیگر نشان می 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 نرخ خطا را هرچقدر که می 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 کلمه 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. این حرف به این معناست که با احتمال یکنواخت این نقاط را انتخاب می کنیم. بنابراین نحوه کدکردن ما به



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

همین سادگی است، یعنی به جای تعریف صریح کد (مثلا نوشتن کد خطی با یک ماتریس مولد یا ماتریس پاریته و نظایرآن) تنها2^k نقطه را به طور تصادفی در این فضا اختیار می کنیم و این نقاط را به عنوان کد کلمه 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 می نامیم و آن را باC_1 نشان می دهیم. دقت کنید که ممکن است بعضی از کلمات کد به هم نزدیک انتخاب شوند که در این صورت احتمال رخ دادن خطا برای این کد کلمه ها زیاد است. در عوض بعضی فاصله یک کد کلمه می تواند از بقیه خیلی زیاد باشد که در این صورت احتمال رخ دادن خطا یعنی تبدیل این کد کلمه به یک کد کلمه 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 C_1، C_2 و نظایر آن نشان می دهیم. حال یک کد مشخص تصادفی را در نظر می گیریم. پس از این نحوه کد کردن می بایست به نحوه تصحیح خطا بپردازیم. نحوه اصلاح خطا نیز خیلی ساده است. هر کلمه ای که دریافت می کنیم به دورش یک کره به شعاع $(H(q)+\delta)$ رسم می کنیم. اصطلاحاً این به این معناست که مجموعه تمام نقاطی را در نظر می گیریم که دارای $2^{n(H(q)+\delta)}$ در بر داشته باشند. هرگاه یک کد کلمه درون این کره قرار داشت می گوییم کلمه ای که ارسال شده است همان نقطه بوده است و کلمه دریافت شده را به همان کلمه اصلاح می کنیم، شکل ؟؟

.

■ تمرین: فرض کنید که $n = 20$ ، و$k = 10$است. احتمال اینکه در یک کد تصادفی یک کدکلمه معین انتخاب شود چقدر است؟ احتمال اینکه دو کد کلمه که فاصله هامینگ آنها یک باشد انتخاب شوند چقدر است؟ فرض کنیدکه $q = 1/4$باشد. در این صورت یک کره به شعاع $H(q)$ را در نظر بگیرید. تعداد نقاط درون این کره چند تاست. احتمال اینکه هیچ نقطه ای درون این کره نباشد چقدراست؟ احتمال این که فقط یک نقطه درون این کره باشد چقدر است؟ احتمال اینکه فقط دو نقطه درون این کره باشد چقدر است؟

همانطور که مثال بالا نشان می دهد احتمال اینکه این کره شامل هیچ کد کلمه ای نباشد خیلی کم است. به طور معمول این کره شامل یک نقطه است که همان کدکلمه اولیه ای است که مخابره شده است خطا وقتی رخ می دهد که این کره شامل یک کد کلمه دیگر باشد . احتمال این خط را می توانیم محاسبه کنیم. احتمال این که یک نقطه دیگر به طور اتفاقی درون این کره قرار گرفته باشد برابراست با نسبت حجم این کره به حجم کل فضا یعنی

$$P_1 = \frac{2^{n(H(q)+\delta)}}{2^n}. \qquad (۵)$$

از آنجا که تعداد نقاط برابر است با $2^k = 2^{nR}$ ، احتمال اینکه یک نقطه درون این کره قرار بگیرد برابر است با

$$P_{error} = 2^{nR} P_1 = 2^{nR} \frac{2^{n(H(q)+\delta)}}{2^n} = 2^{n(R-C+\delta)}. \qquad (۶)$$

این رابطه نشان می دهد که با انتخاب مناسب n می توانیم نرخ مبادله اطلاعات یعنی R را هرچقدر که می توانیم به C نزدیک کنیم و درعین حال خطا را هرچقدر که می خواهیم کم کنیم.

■ تمرین:اگر بخواهیم که میزان خطا کمتر از 0.001 شود و$R \leq C - 0.01$باشد طول کد کلمه ها حداقل چقدر باید باشد؟ اگر بخواهیم با همین نرخ اطلاعات را مبادله کنیم ولی احتمال خطا را به زیر 0.00001 برسانیم حداقل طول کد کلمه 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 کلمه های درون یک کد متوسط گرفته ایم و باردیگر روی تمام کدهای تصادفی $C_1, C_2, \cdots$ متوسط گرفته ایم. در واقع عبارت خطای P_{error} را می بایست با $\langle\overline{P_{error}}\rangle$ عوض کنیم. به طور دقیق تر ثابت کرده ایم که می توان با انتخاب طول مناسب برای کد کلمه ها نرخ مبادله اطلاعات را هرچقدر که می خواهیم به ظرفیت کانال نزدیک کنیم آنچنان که خطا از هر مقداری که می خواهیم کمتر باشد یعنی

$$\langle\overline{P_{error}}\rangle \leq \epsilon, \tag{۷}$$

که در آن

$$\langle O\rangle = \frac{1}{\mathcal{Z}}\sum_{C} O(C), \tag{۸}$$

یک متوسط روی تمام کدهای تصادفی C است. در این جا $\mathcal{Z}$ تعداد کدهای تصادفی است. هم چنین

$$\overline{O} = \frac{1}{2^{nR}}\sum_{i=1}^{2^{nR}} O(i), \tag{۹}$$

که در آن متوسط روی تمام کدکلمه های درون یک کد محاسبه می شود. حال از رابطه ۷ نتیجه می شود که حتما یک کد تصادفی مثل C وجود دارد که برای آن رابطه زیر برقرار است:

$$\overline{P_{error}}(C) \leq \epsilon. \tag{۱۰}$$

اما این عبارت متوسط خطا را روی تمام کدکلمه ها نشان می دهد. ممکن است برای بعضی از کد کلمه ها خطا بیشتر از ϵ و برای بعضی دیگر کمتر از ϵباشد. این کد البته مطلوب ما نیست چرا که ما کدی می خواهیم که احتمال خطا برای تمام کدکلمه هایش از ϵ کمتر باشد. برای رسیدن به این کد تنها کافی است که کد را کمی خلوت کنیم و بعضی از کد کلمه ها را حذف کنیم. تنها یک نگرانی وجود دارد و آن اینکه با خلوت کردن کد و دور ریختن این نوع کلمات تعداد خیلی کمی کد کلمه 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 کلمه ها (یا رشته ها) می توانیم به کدی برسیم که خطا برای تک تک کد-کلمه 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. خطای مربوط به کد کلمه اول را با E_1 ، خطای مربوط به کد کلمه دوم را با E_2 نشان می دهیم و همینطور تا آخر. در این صورت می دانیم که متوسط این خطا ها از ϵ کمتر است. یعنی

$$\langle E \rangle = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} E_i \leq \epsilon. \qquad (۱۱)$$

در این صورت بنا بر لم چبیشف می دانیم که احتمال اینکه یک خطا از 2ϵ بیشتر باشد در نامساوی زیر صدق می کند:

$$P(E \geq 2\epsilon) \leq \frac{\langle E \rangle}{2\epsilon} \leq \frac{\epsilon}{2\epsilon} = \frac{1}{2}. \qquad (۱۲)$$

معنای این حرف این است که کمتر از نصف تعداد کل کد کلمه ها مقدار خطایشان 2ϵ بیشتر است و اگر این تعداد از کد کلمه ها را دور بریزیم بقیه همه خطایشان از 2ϵ کمتر است و این همان چیزی است که می خواستیم ثابت کنیم.

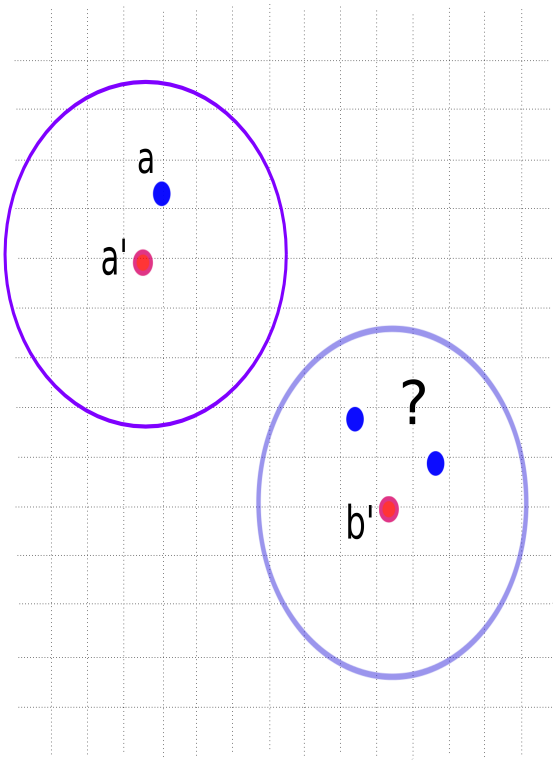
بنابراین با دور ریختن کمتر از نیمی از کد کلمه ها به کدی می رسیم که خطا برای تمام کد کلمه هایش از 2ϵ کمتر است. کد جدیدی که ساخته ایم تعداد کد کلمه هایش برابر است با:

$$\frac{2^{nR}}{2} = 2^{n(R-\frac{1}{n})} \qquad (۱۳)$$

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.

۳ تعریف ظرفیت درحالت کلی

دربخش قبل خود را به منبع پیامی محدودکردیم که دارای آنتروپی بیشینه بود زیرا دراین منبع حرف 0 وحرف 1 هردو بااحتمال $\frac{1}{2}$ رخ می دادند وبنابراین داشتیم $H(X) = H(\frac{1}{2}) = 1$. دراین بخش حالت کلی را درنظرمی گیریم که درآن آنتروپی منبع دلخواه است . فرض می کنیم که منبع حرف 0 را با احتمال p و حرف 1 را بااحتمال $1-p$ تولید کند. درنتیجه آنتروپی منبع



شکل ۲: به دور هر رشته دریافتی یک کره با شعاع مناسب رسم می کنیم. هرگاه یک کدکلمه درون این کره وجود داشته باشد می گوییم همان کلمه ارسال شده و در اثر خطا به این رشته دریافتی تبدیل شده است. خطا وقتی رخ می دهد که وقتی چنین کره ای حول یک رشته دریافتی رسم می کنیم شامل یک کد کلمه دیگر باشد.

برابراست با

$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}. \qquad (۱۴)$$

باید دقت کنیم که در اینجا منظور ما از منبع همان رشته صفر و یک هایی است که پس از کدکردن الفبا بوجود آمده اند و این آنتروپی با آنتروپی الفبا فرق دارد. تمرین زیر این موضوع را روشن می کند.

در ابتدای بحث های خود وقتی که می خواستیم یک روش کد گذاری مشخص را برای کد کردن k بیت در n بیت تعریف می کنیم نقطه شروع ما رشته های k بیتی بود که آنها را به طریق معینی کد می کردیم. در این جا چون هدف ما پیداکردن ظرفیت کانال و اثبات وجود یک کدگذاری است که این ظرفیت را اشباع کند روش استدلال ما اندکی متفاوت است اگر چه از

نظر مفهومی با استدلال هایی که قبلا انجام داده ایم معادل است.

رشته های n حرفی از منبع $X = \{0,1\}$ را در نظر می گیریم. هرگاه یک رشته n حرفی انتخاب کنیم با احتمال نزدیک به یک این رشته یک رشته نمونه خواهد بود و هرچه که n زیادتر باشد این احتمال نیز به یک نزدیک تراست. تعداد این رشته ها برابراست با $2^{nH(X)}$. می دانیم که کانال کلاسیک دارای یک نوفه است و بنابراین همه این رشته ها را نمی توانیم به سلامت از کانال عبور بدهیم زیرا این رشته ها خیلی نزدیک به هم هستند و براحتی باهم اشتباه می شوند. چاره کار آن است که تنها تعدادی از این رشته ها را انتخاب کرده و آنها را به عنوان رشته های کد شده از کانال عبور دهیم به نحوی که بتوانیم در مقصد خطای اعمال شده روی آنها را اصلاح کنیم و رشته های صحیح را بازیابی کنیم. هرگاه بتوانیم 2^k رشته را در این فضا انتخاب کنیم که دارای خصلت فوق باشند می گوییم دراین کد k بیت را در n بیت کد کرده ایم. بنابراین بازهم به ترتیب قبل پیش می رویم به این نحو که به طور تصادفی 2^k نقطه را انتخاب می کنیم و این کد ما را تعریف می کند. برای اصلاح خطا به دور هر کلمه یا رشته ای مثل y که دریافت می کنیم یک کره به شعاع $H(X|y)+\delta$ رسم می کنیم. این کره دارای $2^{n(H(X|Y)+\delta)}$ رشته یا کلمه را در بر می گیرد. این رشته ها محتمل ترین رشته هایی x ای هستند که می توانسته اند در اثر عبور از کانال و تحمل خطا به رشته ی دریافت شده ی y تبدیل شده باشند. رجوع کنید به شکل ۱.۲ .

این موضوع را بعداً نشان می دهیم. احتمال خطا این است که در درون این کره علاوه بر رشته اولیه x یک رشته دیگر نیز وجود داشته باشد. یعنی این که یک رشته دیگر در اثر خطای زیاد از جای خود حرکت کرده و درون این کره افتاده باشد. این خطا را براحتی می توان حساب کرد . از آنجا که کد تصادفی است احتمال این که یک رشته معین درون این کره قرار گرفته باشد برابر است با:

$$P_1 = \frac{2^{n(H(X|Y)+\delta)}}{2^{nH(X)}}. \qquad (۱۵)$$

این احتمال این است که یک رشته معین در اثر خطا درون این کره قرار گرفته باشد. احتمال این که هر کدام از رشته های کد درون این کره قرار گرفته باشند برابر است با:

$$P_{error} = 2^k \times P_1 = 2^{nR} \times \frac{2^{n(H(X|Y)+\delta)}}{2^{nH(X)}} = 2^{n(R-H(X)+H(X|Y)+\delta)}. \qquad (۱۶)$$

بنابراین هرگاه که

$$R > H(X) - H(X|Y) \qquad (۱۷)$$

باشد می توانیم با بزرگ تر کردن طول رشته ها خطا را هر قدر که می خواهیم کوچک کنیم. به این ترتیب به تعریف زیر برای ظرفیت یک کانال می رسیم:

$$C = max_X I(X;Y) = max_X \left(H(X) - H(X|Y)\right). \qquad (۱۸)$$

ودرنتیجه نشان دادیم که امکان مخابره اطلاعات باهرنرخ R که کمترازظرفیت کانال باشد امکان پذیراست.

■ تمرین: فرض کنید که $H(X) = 1/3$. کانال را نیز به صورت زیر تعریف می کنیم:

$$P(0|1) = P(1|0) = p. \qquad (۱۹)$$

الف: عبارت $H(X) - H(X|Y)$ را حساب کنید.

ب: اگر بخواهیم که نرخ مبادله اطلاعات به اندازه $\frac{1}{100}$ از این مقدار کمتر باشد و درعین حال خطا از 0.0001 کمتر باشد طول رشته ها از چه عددی می بایست بیشتر باشد؟

حال به آخرین قسمت استدلال می رسیم و آن اینکه به طور متوسط تعداد رشته های x ای که می توانسته اند به یک y تبدیل شده باشند برابراست با $2^{nH(X|Y)}$. برای فهم این موضوع دقت می کنیم که تعداد رشته های نمونه از نوع x برابر است با $2^{nH(X)}$. تعداد جفت رشته های نمونه های (x,y) برابر است با $2^{nH(X,Y)}$. هم چنین تعداد رشته های نمونه از نوع y برابر است با $2^{nH(Y)}$. مطابق با قضایایی که در درس پیشین ثابت کردیم تمام این رشته ها با احتمال یکنواخت توزیع شده اند. بنابراین از تعداد $2^{nH(X,Y)}$ رشته نمونه با احتمال $2^{-nH(Y)}$ به یک رشته برمی خوریم که یک y معین داشته باشد. در نتیجه تعداد جفت رشته های نمونه که y آنها یک رشته معین باشد برابراست با

$$2^{-nH(Y)} \times 2^{nH(X,Y)} = 2^{nH(X|Y)}. \qquad (۲۰)$$

واین همان چیزی بود که می خواستیم ثابت کنیم. به این ترتیب اثبات رابطه اصلی در مورد ظرفیت کانال های کلاسیک و هم چنین درس مربوط به اطلاعات کلاسیک به پایان می رسد.