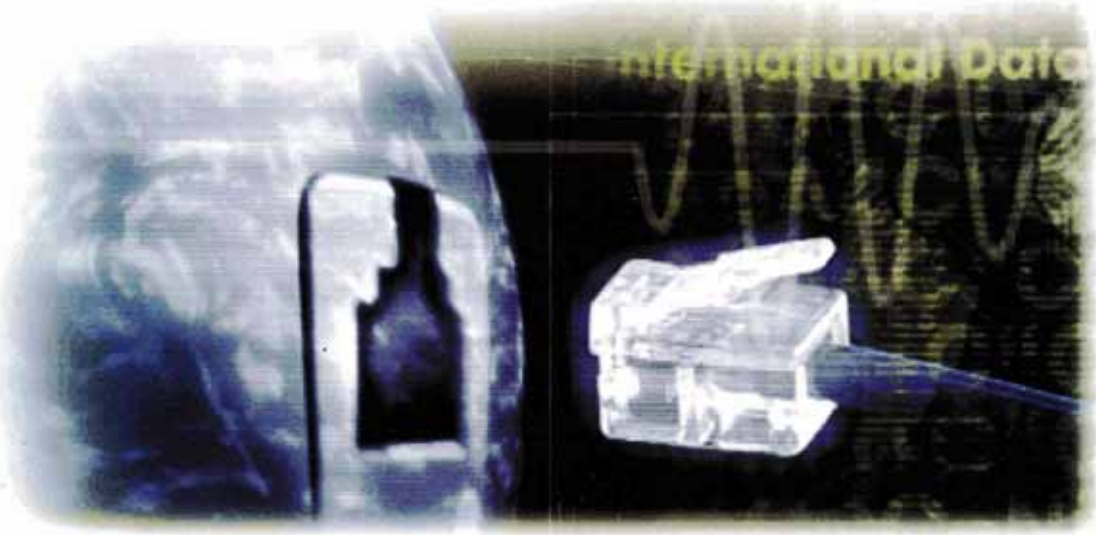
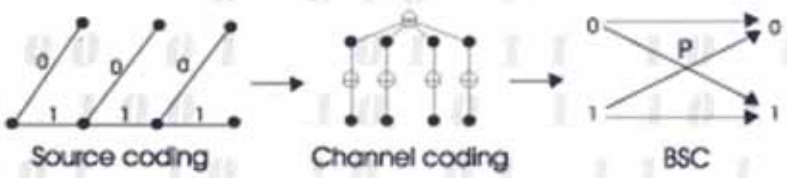
دانشگاه صنعتی اصفهان
مرکز نشر

نظریه اطلاعات و کدگذاری

تألیف: ج. آ. جانز و ج. م. جانز
ترجمه: دکتر مرتضی اسماعیلی





شابک : ۳-۹۷-۶۰۲۹-۹۶۴

نظریه اطلاعات و کدگذاری

تألیف:
گ. آ. جانز – ج. م. جانز

ترجمه:
دکتر مرتضی اسماعیلی
عضو هیئت علمی دانشگاه صنعتی اصفهان
۱۳۸۳

دانشگاه صنعتی اصفهان
مرکز نشر

شماره کتاب ۸۶ گروه علوم ۳۰

نظریه اطلاعات و کدگذاری

مؤلفین ..... :	گ. آ. جانز و ج. م. جانز
ترجمه ..... :	دکتر مرتضی اسماعیلی
حروف‌چینی کامپیوتری ..... :	مرکز نشر دانشگاه صنعتی اصفهان
لیتوگرافی، چاپ و صحافی ..... :	مرکز نشر دانشگاه صنعتی اصفهان
ناشر ..... :	مرکز نشر دانشگاه صنعتی اصفهان
چاپ اول ..... :	بهار ۸۳
تیراژ ..... :	۱۵۰۰ جلد
شابک ..... :	۳-۹۷-۶۰۲۹-۹۶۴
قیمت ..... :	۲۰۰۰ تومان

جانز، گارت، ۱۹۴۶- م. Jones, Gareth A.

نظریه اطلاعات و کدگذاری/ تالیف گ. آ. جانز ج. م. جانز؛ ترجمه مرتضی اسماعیلی.ـــ اصفهان: مرکز نشر دانشگاه صنعتی اصفهان، ۱۳۸۳.

IX، ۲۱۱ ص.

ISBN 964-6029-97-3

فهرست‌نویسی بر اساس اطلاعات فیپا.

عنوان اصلی: Information and coding theory.

واژه‌نامه.

۱. نظریه اطلاعات. ۲. نظریه کدگذاری.

الف. جانز، جوزفین مری، ۱۹۴۶ - م. Jones, Josephine Mary ب. اسماعیلی، مرتضی، ۱۳۳۶- ،مترجم. ج. مرکز نشر دانشگاه صنعتی اصفهان.

د. عنوان.

۹ت۹۴ج/ Q۳۶۰ ۱۳۸۲ ۰۰۳/۵۴

کتابخانه ملی ایران ۸۲-۳۰۱۶۶م

پیش‌گفتار مترجم

نظریه اطلاعات و کدگذاری دانش انتقال دقیق و اقتصادی داده‌ها از نقطه‌ای به نقطه دیگر یا از زمانی به زمان دیگر است. به‌عنوان نمونه می‌توان به مکالمات تلفنی و ارسال تصویر از دیگر سیارات به زمین توسط ایستگاه‌های فضایی و نگهداری داده‌ها روی CD اشاره نمود.

در حالی‌که نظریه اطلاعات توجه به روش‌های کدگذاری خروجی یک متغیر تصادفی با استفاده بهینه از توزیع احتمال آن متغیر به‌منظور کاهش افزونگی دارد، در نظریه کدگذاری کانال هدف اضافه نمودن افزونگی به‌منظور ایجاد توان‌مندی لازم جهت خنثی‌نمودن توان تخریبی پارازیت روی داده‌ها است. از این‌رو در اولی تا حد امکان افزونگی کاهش یافته و در دومی به‌قدر کافی افزونگی به داده‌ها افزوده می‌شود. در هر دو مرحله هدف ایجاد بستر مناسب جهت ارسال سریع، درست و ارزان داده‌ها از فرستنده به گیرنده است.

کتاب حاضر حاوی مفاهیم و روش‌های اساسی کدگذاری منبع (خروجی یک متغیر تصادفی) و کانال است. با توجه به ساختار ارائه مطالب، می‌توان از این کتاب به‌عنوان منبع اصلی ارائه یک درس کدگذاری (منبع و کانال) در مقطع کارشناسی برای رشته‌های ریاضی کاربردی، کامپیوتر، و مخابرات استفاده نمود.

تلاش شده است که واژه‌ها و نمادها کاملاً به فارسی بیان شوند. با این حال به‌جهت وجود محدودیت‌ها دو مورد استثنا رخ داده است. براساس ساختار متن و جمله از هر دو کلمه 'کدگشایی' و ' دکد کردن' استفاده شده است. به‌منظور یکسان نمودن بزرگی اعداد موجود در یک رابطه یا جدول نماد '/' به‌عنوان علامت تقسیم به‌کار رفته و از این‌رو بر خلاف متون فارسی برای ممیز از '.' استفاده شده است؛ بنابراین به‌عنوان نمونه اعداد ۴/۹ و ۵.۳ به‌ترتیب چهار تقسیم بر نه و پنج و سه‌دهم خوانده می‌شوند.

مرتضی اسماعیلی
عضو هیئت علمی دانشگاه صنعتی اصفهان
بهمن ۱۳۸۲

پیش‌گفتار مؤلف

این پیش‌گفتار در زمانی در حال نوشته شدن است که قرن بیستم در حال پایان است. تاریخ نویسان احتمالاً از این قرن به‌عنوان قرن اطلاعات یاد خواهند کرد، درست مانند قرن قبل که به مرحله صنعتی شدن نسبت داده می‌شود. پیشرفت‌های فنی متوالی مانند تلفن، رادیو، تلویزیون، کامپیوتر و اینترنت تأثیرات عمیقی روی روش زندگی ما داشته‌اند. می‌توانیم تصویرهایی از سطح مریخ یا شکل ابتدایی جهان را ببینیم. محتوای کتاب‌های یک قفسه از کتابخانه قابل فشرده شدن روی یک قطعه پلاستیکی تقریباً بی‌وزن است. میلیاردها نفر می‌توانند یک بازی فوتبال را هم‌زمان تماشا کنند، یا موفق به برقراری تماس فوری با دوستان خود در سراسر دنیا بدون ترک خانه باشند. به‌طور خلاصه، در حال حاضر می‌توان انبوهی از اطلاعات را با سرعت، دقت و کارایی اعجاب‌انگیزی ذخیره، ارسال و پردازش نمود.

واضح است که این پیشرفت‌ها بدون مبانی نظری اتفاق نمی‌افتد و همچون در بسیاری موارد دیگر بخش عمده‌ای از این پیشرفت مرهون ریاضیات است. بسیاری از پیشرفت‌های اولیه ریاضی در این زمینه در اواسط قرن بیستم به‌وسیله مهندسین حاصل شد و این اغلب با اتکا بر درک شهودی و تجربی بوده تا این‌که یک دانش نظری عمیق آنها را به کشف‌هایشان هدایت کرده باشد. ریاضی‌دانان، که از دیدن کاربردهای جدید موضوع کاری خود خوشحال بودند خیلی زود وارد صحنه شدند و مثال‌های عملی مهندسین را به نظریه‌های وسیعی همراه با تعاریف، قضایا و اثبات‌ها توسعه دادند. شاخه‌های جدیدی از ریاضیات خلق شدند و چندین شاخه قدیمی‌تر تحت تأثیر کاربردهای غیرمنتظره تقویت شدند: چه کسی می‌توانست پیش‌بینی کند که کدهای تصحیح کننده خطا می‌توانند بر مبنای خم‌های جبری روی میدان‌های متناهی باشند یا این‌که سیستم‌های رمزنگاری وابسته به اعداد اول باشند؟

نظریه اطلاعات و نظریه کدگذاری دو جنبه مرتبط با هم از مسئله چگونگی ارسال سریع و دقیق اطلاعات از یک منبع از طریق یک کانال به یک گیرنده است. این در برگیرنده مسئله چگونگی ذخیره کردن اطلاعات نیز می‌شود که در آن گیرنده می‌تواند همان منبع ولی در زمان بعد باشد. به‌عنوان مثال، کشف فضا منجر به ایجاد تقاضا برای ارسال دقیق یک سیگنال خیلی ضعیف از طریق یک کانال فوق‌العاده شلوغ شده است؛ دلیلی برای فرستادن یک فضاپیما به مریخ وجود ندارد اگر نتوان پیام‌هایی را که او می‌فرستد شنید و کدگشایی کرد.

این نظریه در ساده‌ترین فرم خود از روش‌های مقدماتی نظریه احتمال و جبر خطی استفاده می‌کند، اگرچه پیشرفت‌های بعدی بر مبنای موضوعاتی چون ترکیبیات و هندسه جبری بوده است.

یک مسئله مهم چگونگی فشرده‌سازی اطلاعات به‌منظور ارسال سریع و یا ذخیره اقتصادی آن می‌باشد. این‌را می‌توان با کاهش افزونگی انجام داد: یک مثال آشنا استفاده از اختصاراتی چون 'UK' ، 'IBM' و 'radar' به‌جای نام کامل می‌باشد که بسیاری از سمبل‌های آن‌ها از نقطه‌نظر محتوای اطلاعاتی زاید هستند. مشابهاً، ما اغلب نام نزدیک‌ترین دوستان و بستگان خود را کوتاه به‌کار می‌بریم، به‌طوری‌که William تبدیل به Will یا Bill می‌شود.

یک مسئله مهم دیگر چگونگی کشف و اصلاح خطا در اطلاعات است. نمی‌توان همیشه به انسان و ماشین از جهت عدم ارتکاب اشتباه اطمینان کرد، و اگر این اشتباهات تصحیح نشوند پیامدهای نامطلوبی را می‌توانند در پی داشته باشند. در این‌جا شیوه حل مسئله افزایش افزونگی است که با اضافه کردن سمبل‌هایی که باعث تقویت و حفاظت پیام می‌شود صورت می‌گیرد. از این‌رو الفبای NATO متشکل از Alpha، Bravo، Charlie، ...، که در ارتباطات گفتاری توسط نیروهای مسلح، خطوط هوایی و سرویس‌های اضطراری به‌کار می‌رود حروف A،B،C،...، را با کلماتی جایگزین می‌کند که به‌نحوی انتخاب شده‌اند که تا حد امکان به لحاظ شنیداری متفاوت باشند؛ به‌عنوان نمونه حروف B و V اغلب از نظر صدا غیرقابل تمیز هستند (اساساً این دو در بعضی زبان‌ها یکی هستند)، اما ممکن نیست که Victor به‌عنوان Bravo درنظر گرفته شود حتی اگر به‌صورت Bictor شنیده شود.

نظریه اطلاعات که بخش عمده آن ریشه در مقاله مهم سال ۱۹۴۸ شانون [Sh 48] دارد، از توزیع احتمال برای اندازه‌گیری اطلاعات (از طریق تابع آنتروپی)، و ارتباط دادن آن با میانگین طول کلمه در کدگذاری‌های آن اطلاعات استفاده می‌کند. بالاخص، قضیه اساسی شانون وجود کدهای خوب تصحیح کننده خطا را تضمین کرده، و هدف نظریه کدگذاری استفاده از روش‌های ریاضی برای ساخت یک چنین کدهایی به‌همراه الگوریتم‌های کارا برای استفاده از آنها می‌باشد. نظریه کدگذاری، علی‌رغم نام آن، درگیر در مطالعه کدهای مخفی نیست: این موضوع که رمزنگاری می‌باشد ارتباط نزدیکی با نظریه اطلاعات از طریق مفاهیم آنتروپی و افزونگی دارد، ولی ما به جهت این‌که فنون ریاضی مربوطه نسبتاً متفاوت هستند در این‌جا به آن نمی‌پردازیم.

این کتاب بر مبنای یک درس سال سوم دوره کارشناسی که در اوایل دهه ۱۹۸۰ در دانشگاه سوتامتن[1] ارائه شد تنظیم شده و هدف آن تلاشی برای توصیف ایده‌های اصلی نظریه اطلاعات و نظریه کدگذاری می‌باشد. پیش‌نیازهای اصلی آن نظریه مقدماتی احتمال و جبر خطی و مقدار کمی حسابان است. اکثر کتاب‌های درسی در این زمینه تا حد زیادی و یا به‌طور کامل تنها روی یکی از دو موضوع نظریه اطلاعات و نظریه کدگذاری تمرکز دارند. با این حال، این دو موضوع از طریق قضیه شانون تا حد زیادی به یکدیگر مرتبط بوده، و ما

[1] Southampton University

احساس می‌کنیم که دلایل قوی برای فراگیری توأم آنها، حداقل در مراحل اولیه، وجود دارد.

فصل‌های ۱ — ۵ (حدود ۶۰٪ کتاب) اساساً راجع به نظریه اطلاعات است. فصل اول، که پیش‌نیاز خیلی کمی دارد، نشان می‌دهد که چگونه اطلاعات را باید کدگذاری کرد به‌قسمی‌که عمل کدگشایی غیرمبهم و لحظه‌ای باشد. نتایج اصلی این فصل قضیه سردناس — پترسن[۱] و نامساوی‌های کرفت و مک‌میلان[۲] می‌باشند که در رابطه با وجود یک چنین کدهایی هستند. فصل ۲ به معرفی کدهای هافمن می‌پردازد که، نسبتاً مانند کد مرس[۳]، میانگین طول کلمه را با نسبت دادن کدکلمه‌های کوتاه‌تر به سمبل‌های محتمل‌تر می‌نیمم می‌کند؛ در این‌جا (مانند فصل‌های ۳—۵) از نظریه مقدماتی احتمال، و در واقع توزیع احتمال متناهی استفاده می‌کنیم. در فصل ۳، از تابع آنتروپی که بر پایه یک توزیع احتمال و لگاریتم آنها است در جهت اندازه‌گیری اطلاعات و مرتبط نمودن آن، به‌وسیله قضیه‌ای از شانون، به میانگین طول کلمات کدگذاری‌ها استفاده می‌کنیم. فصل ۴ به مطالعه چگونگی ارسال اطلاعات از طریق یک کانال، که ممکن است به دلیل وجود پارازیت خطاهایی رخ دهد، اختصاص دارد؛ احتمال‌های شرطی امکان تعریف یک دستگاه آنتروپی را فراهم می‌سازد که بر مبنای آن می‌توان اطلاعات را از چند نقطه‌نظر، مثلاً از دیدگاه فرستنده و گیرنده، اندازه‌گیری کرد. این منجر به مفهوم ظرفیت کانال می‌شود که عبارت است از بیشترین مقدار اطلاعات که یک کانال می‌تواند ارسال کند. در فصل ۵ قضیه اساسی شانون را ملاحظه می‌کنیم که می‌گوید علی‌رغم وجود پارازیت، می‌توان اطلاعات را با دقت دلخواه و با نرخ به‌قدر دلخواه نزدیک به ظرفیت کانال ارسال نمود. رئوس اثبات این قضیه را برای کانال ساده ولی مهم دوتایی متقارن ارائه داده و سپس اثبات کامل قضیه را برای این کانال در ضمیمه C می‌آوریم؛ این اثبات متکی بر تنها نتیجه پیشرفته مورد نیاز ما از نظریه احتمال، یعنی قانون اعداد بزرگ، است که در ضمیمه B ارائه شده است.

ایده اساسی قضیه شانون این است که می‌توان اطلاعات را با صحت بالائی ارسال کرد و این کار با استفاده از کدکلماتی که به‌قدر کافی با هم تفاوت دارند، و از این‌رو حتی اگر بعضی از سمبل‌های آنها نادرست دریافت شوند گیرنده آنها را با هم اشتباه نخواهد کرد، صورت می‌گیرد (Bravo و Victor را به یاد آورید). متأسفانه قضیه و اثبات آن چگونگی پیدا کردن مثال‌های مشخصی از این کدها را بیان نمی‌کنند، و این کار هدف نظریه کدگذاری است که در فصل‌های ۶ و ۷ به آن می‌پردازیم. در این فصل‌ها که نسبتاً از فصل‌های قبلی خود طولانی‌تر هستند چند مثال نسبتاً ساده از کدهای تصحیح کننده خطا را معرفی می‌کنیم. برای این منظور در فصل ۶ از روش‌های مستقیم و مقدماتی استفاده می‌کنیم؛ نتیجه اصلی این فصل کران کره‌چینی همینگ[۴] است که با به‌کار بردن یک ایده ساده هندسی یک کران بالا روی تعداد کدکلماتی که بتوانند تعداد خطای مفروضی را تصحیح کنند ارائه می‌دهد.

در فصل ۷ مثال‌های نسبتاً پیشرفته‌تری از کدهای تصحیح کننده خطا را می‌سازیم؛ این کار

---

[۱] Sardinas - Patterson theorem
[۲] Kraft and Mcmillan inequalities
[۳] Morse code
[۴] Hamming's sphere - packing bound

با استفاده از جبر خطی و نظریه ماتریس‌ها، و مشخصاً مفاهیمی چون فضاهای برداری، زیر فضاها، مبنا و بعد، رتبه ماتریس و عملیات سطری و ستونی صورت می‌گیرد. همچنین به‌صورت مختصر چگونگی ارتباط ایده‌هایی از ترکیبیات و هندسه، مانند طرح‌های قالبی و هندسه‌های تصویری، را با کدها نشان می‌دهیم.

محدودیت‌های معمول فضا و زمان ما را وادار به حذف موضوعات جالبی چون رابطه با رمزنگاری که در بالا به آن اشاره شد، و اشاره مختصر به چند مطلب دیگر نموده است. به‌عنوان نمونه، در نظریه اطلاعات منابع مارکف (آنهایی که حافظه‌ای از پیشامدهای قبل دارند) تنها به‌عنوان یک تمرین ظاهر شده، و مشابهاً در نظریه کدگذاری به کدهای دوری و ارتباط آنها با حلقه چندجمله‌ای‌ها اشاره نشده است. در عوض در انتهای کتاب پیشنهادهایی را برای مطالعه بیشتر آورده‌ایم.

یک درس نظریه اطلاعات را می‌توان براساس فصل‌های ۱—۵، و احتمالاً با مطالب بیشتری راجع به منابع مارکف و یا رابطه با رمزنگاری تنظیم نمود. یک درس نظریه کدگذاری می‌تواند بر پایه فصل‌های ٦ و ۷ و مقداری پیش‌نیاز از فصل ۵ به همراه مطالب اضافی دیگری همچون کدهای دوری یا شمارش وزن استوار باشد.

تلاش کرده‌ایم تا اعتبار لازم را به بنیان‌گزاران ایده‌های اصلی معرفی شده در کتاب داده، و از منابع اولیه نتایج، مثال‌ها، و تمرینات قدردانی کنیم. بدون شک از این دیدگاه بدون ضعف نبوده ولی ضعف‌های احتمالی بدون هیچ قصد و عمدی هستند.

فهرست مندرجات

فصل ۱

کدگذاری منبع

کلمات، کلمات، کلمات.(Hamlet)

در این فصل چگونگی کدگذاری اطلاعات ظاهر شده از یک منبع به‌قسمی‌که بعداً بدون تأخیر و ابهام قابل دکد شدن باشد بررسی می‌شود. این دو شرط منجر به مفاهیم کدهای لحظه‌ای[۱] و کدهای یکتا دکدپذیر[۲] می‌شوند. شرط‌های لازم و کافی برای این‌که یک کد دارای چنین خواصی باشد ارائه شده و نحوه ساخت این کدها بررسی می‌شود. نامساوی‌های کرفت[۳] و مک‌میلان[۴] را ثابت خواهیم نمود که براساس آن‌ها چنین کدهایی موجود هستند اگر و فقط اگر حاوی تعداد کافی از کدکلمه‌های طولانی باشند.

۱.۱ تعاریف و مثال‌ها

نظریه اطلاعات مربوط به انتقال اطلاعات از یک فرستنده به یک گیرنده از طریق یک کانال می‌باشد. فرستنده و گیرنده می‌توانند انسان یا ماشین باشند. در اکثر اوقات این دو متفاوت هستند، ولی وقتی اطلاعات ذخیره می‌شود تا بعداً اعاده گردد گیرنده می‌تواند همان فرستنده در زمان آینده باشد. فرض خواهیم کرد که اطلاعات از یک منبع S می‌آید که دنباله $\mathbf{s} = X_1X_2X_3\ldots$ را تولید می‌کند. به‌عنوان مثال X_n ممکن است n امین سمبل یک پیام بوده و یا نتیجه n امین دفعه از یک آزمایش تکراری باشد. در عمل این دنباله همیشه متناهی خواهد بود (هیچ چیز همیشگی نیست)، از نقطه‌نظر تئوری درنظر گرفتن دنباله‌های نامتناهی نیز گاهی اوقات مفید است. فرض خواهیم کرد که هر سمبل X_n یک عضو

[۱] Instantaneous [۲] Uniquely decodable [۳] Kraft [۴] McMillan

مجموعه ثابت $S = \{s_1, \ldots, s_q\}$ بوده که S الفبای منبع $\mathcal{S}$ نامیده می‌شود. برای سادگی فرض می‌کنیم احتمال این‌که n امین سمبل X_n برابر s_i باشد، $Pr(X_n = s_i)$، تنها بستگی به i داشته و مستقل از n باشد. بنابراین برای $1 \leq i \leq q$ می‌نویسیم

$$Pr(X_n = s_i) = p_i.$$

پس سمبل‌های مختلف ممکن است احتمال‌های مختلف داشته باشند ولی این احتمال‌ها نسبت به زمان ثابت بوده (پس $\mathcal{S}$ پایا است) و بستگی به سمبل‌های قبلی X_m ندارند که $m < n$ (بنابراین $\mathcal{S}$ بدون حافظه است). در یک سطح پیشرفته‌تر این فرض‌ها نیز لحاظ می‌شوند، ولی در این‌جا ما از آن‌ها صرف‌نظر می‌کنیم. مانند هر توزیع احتمال، احتمال‌های p_i نامنفی بوده و در شرط زیر صدق می‌کنند:

$$\sum_{i=1}^{q} p_i = 1. \qquad (1.1)$$

به زبان آماری، $\mathcal{S}$ را می‌توان یک دنباله از متغیرهای تصادفی مستقل X_n با توزیع یکسان و توزیع احتمال (p_i) درنظر گرفت.

مثال ۱.۱ $\mathcal{S}$ یک تاس نااریب بوده و $S = \{1, 2, \ldots, 6\}$. در این‌جا $q = 6$، $s_i = i$ برای $1 \leq i \leq 6$، X_n نتیجه n امین پرتاب بوده، و $p_i = \frac{1}{6}$. یک تاس اریب وضعیتی مشابه دارد با این تفاوت که p_iها یکسان نخواهند بود.

مثال ۲.۱ $\mathcal{S}$ را هوای یک مکان خاص درنظر می‌گیریم. X_n را نیز معرف هوا در روز nام انتخاب می‌کنیم. برای سادگی می‌توان فرض کرد که S شامل سه نوع هوای متمایز (خوب، معتدل، و بد) باشد. پس p_i، $1 \leq i \leq 3$، احتمال هر یک از این سه نوع هوا است، مثلاً $p_1 = \frac{1}{4}$، $p_2 = \frac{1}{2}$، $p_3 = \frac{1}{4}$. (در این‌جا تغییرات فصلی را که باعث ایجاد تغییر در توزیع احتمال p_i می‌شود درنظر نمی‌گیریم.)

مثال ۳.۱ $\mathcal{S}$ یک کتاب بوده و S شامل تمامی سمبل‌های به‌کار رفته (حروف، علائم نقطه‌گذاری، اعداد، و غیره) در آن است. X_n معرف n امین سمبل در کتاب بوده و p_i فراوانی i امین سمبل در الفبای منبع می‌باشد. (در این‌جا تأثیر سمبل‌های قبلی را روی احتمال‌ها درنظر نمی‌گیریم؛ به‌عنوان مثال در یک متن انگلیسی همواره بعد از حرف q حرف u ظاهر می‌شود.)

برای کدگذاری یک منبع از یک الفبای متناهی کد، $T = \{t_1, \ldots, t_r\}$، شامل r سمبل t_j استفاده می‌کنیم. در حالت کلی این الفبا متمایز از الفبای منبع، یعنی $S = \{s_1, \ldots, s_q\}$،

است زیرا این الفبا به تکنولوژی کانال بستگی داشته و ارتباطی با منبع ندارد. عدد r را ریشه[۱] و کد را یک کد r–تایی می‌نامیم. در بسیاری از مثال‌ها r برابر ۲ بوده و کد را دوتایی می‌نامند. اکثر کدهای دوتایی، مانند ASCII (که در محاسبه به‌کار می‌رود)، دارای $T = \mathbf{Z}_2 = \{0, 1\}$ هستند. کدهای با $r = 3$ را سه‌تایی می‌نامند. برای کدکردن $\mathcal{S}$، هر سمبل $s_i \in S$ با یک کدکلمه w_i نمایش داده می‌شود که w_i یک دنباله متناهی از سمبل‌های کد می‌باشد. برای کدکردن $\mathbf{s} = X_1X_2X_3\ldots$ هر سمبل $X_n = s_i$ با w_i نمایش داده شده و این یک دنباله $\mathbf{t}$ از سمبل‌های T ایجاد می‌کند. جهت اختصار، کلمات واقع در $\mathbf{t}$ را با علائم نقطه‌گذاری یا فاصله از هم جدا نمی‌کنیم. اگر چنین علائمی به‌کار روند باید با سمبل‌های T نمایش داده شده و در ابتدا یا انتهای هر w_i ظاهر شوند. بنابراین کد مورس[۲] که دوتایی به‌نظر می‌رسد در واقع یک کد سه‌تایی است: سمبل‌ها عبارت هستند از •، ــ، و یک فاصله.

مثال ۴.۱ همچون مثال ۱.۱ فرض کنید $\mathcal{S}$ یک تاس نااریب باشد. قرار دهید $T = \mathbf{Z}_2$ و فرض کنید w_i نمایش دوتایی سمبل s_i، $1 \le i \le 6$، باشد. پس $w_1 = 1$، $w_2 = 10$، ...، $w_6 = 110$. بنابراین یک دنباله از پرتاب‌ها همچون $s = 53214$ به‌صورت $\mathbf{t} = 10111101100$ کد می‌شود.

جهت تشریح بهتر، بعضاً قرارداد خود مبنی بر عدم به‌کارگیری نقطه‌گذاری را شکسته و از علامت پایان کامل یا پرانتز استفاده نموده تا نشان دهیم که $\mathbf{t}$ چگونه به کدکلمه‌های متمایز تجزیه می‌شود. به‌عنوان نمونه در مثال ۴.۱ می‌توان نوشت $\mathbf{t} = 10.11.10.1.100$. این صرفاً برای کمک به خواننده بوده و سمبل‌های نقطه‌گذاری را نباید به‌عنوان بخشی از $\mathbf{t}$ درنظر گرفت.

نیاز به ارائه یک تعریف دقیق‌تری از کد داریم. یک کلمه w در T عبارت از یک دنباله متناهی از سمبل‌های T بوده و طول آن برابر تعداد سمبل‌های به‌کار رفته می‌باشد که با $|w|$ نمایش داده می‌شود. مجموعه تمامی کلمات در T را با T^* نمایش می‌دهیم، این شامل کلمه خالی به‌طول صفر نیز بوده که با ε نمایش داده می‌شود. مجموعه تمامی کلمات غیرخالی در T را با T^+ نمایش می‌دهیم. پس

$$T^* = \bigcup_{n \ge 0} T^n \quad , \quad T^+ = \bigcup_{n > 0} T^n,$$

که T^n مجموعه تمامی کلمات به‌طول n است. یک کد منبع (یا برای سادگی یک کد) $\mathcal{C}$ در واقع یک تابع $S \longrightarrow T^+$ است که در آن s_i به $w_i = \mathcal{C}(s_i)$ نگاشته می‌شود. بسیاری از خواص یک کد تنها به کدکلمه‌های w_i بستگی داشته و تناظر خاص بین آن‌ها و سمبل‌های s_i اهمیتی ندارد. از این‌رو ما اغلب $\mathcal{C}$ را به‌عنوان یک مجموعه متناهی از کلمات $w_1, \ldots, w_q$

[۱] radix [۲] Morse

در T^+ درنظر می‌گیریم. اگر S^* به روشی مشابه T^* تعریف شود، آنگاه C را می‌توان به‌سادگی به یک تابع $S^* \longrightarrow T^*$ تعمیم داد:

$$\mathbf{s} = s_{i_1} s_{i_2} \dots s_{i_n} \longmapsto \mathbf{t} = w_{i_1} w_{i_2} \dots w_{i_n} \in T^*.$$

تصویر این تابع مجموعه C^* است:

$$C^* = \{w_{i_1} w_{i_2} \dots w_{i_n} \in T^* \ : w_{i_j} \in C,\ n \geq 0\}.$$

طول کلمه w_i، یعنی $|w_i|$، را با l_i نمایش می‌دهیم. پس $l_i \geq 1$. متوسط طول در C برابر $L(C)$ تعریف می‌شود:

$$L(C) = \sum_{i=1}^{q} p_i l_i. \qquad (2.1)$$

مثال ۵.۱ کد C معرفی شده در مثال ۴.۱ دارای طول‌های $l_1 = 1$، $l_2 = l_3 = 2$ و $l_4 = l_5 = l_6 = 3$ است، بنابراین

$$L(C) = \frac{1}{6}(1 + 2 + 2 + 3 + 3 + 3) = \frac{7}{3}.$$

هدف ساخت کدهایی چون C است که برای آن‌ها

(الف) یک کدگشایی آسان و غیرمبهم $\mathbf{t} \longmapsto s$ وجود داشته باشد،

(ب) متوسط طول کد، $L(C)$، کوچک باشد.

بقیه این فصل اختصاص به‌خاصیت (الف) داشته، و خاصیت (ب) در فصل بعد بررسی می‌شود.

۲.۱ کدهای یکتا دکدپذیر

یک کد C را یکتا دکدپذیر نامیم هرگاه هر $\mathbf{t} \in T^*$ تحت C حداکثر به یک $s \in S^*$ متناظر شود؛ به‌عبارت دیگر تابع $C : S^* \longrightarrow T^*$ یک‌به‌یک بوده و بنابراین هر $\mathbf{t}$ در تصویر آن به‌طور یکتا دکد شود. فرض بر این است که کدکلمه‌های w_i در C متمایز هستند، برای این‌که اگر $i \neq j$ و $w_i = w_j$، آنگاه $\mathbf{t} = w_i$ می‌تواند به s_i یا s_j دکد شود که این خلاف یکتا دکدپذیر بودن C است. تحت این فرض، تعریف یکتا دکدپذیر بودن C معادل این است که اگر $u_1 u_2 \dots u_m = v_1 v_2 \dots v_n$ در C بوده و $u_1, u_2, \dots, u_m, v_1, v_2, \dots, v_n$ آنگاه $m = n$ و برای هر i تساوی $u_i = v_i$ برقرار است. به زبان جبری می‌گوئیم که هر $\mathbf{t} \in C^*$ به‌طور یکتا به حاصل‌ضرب کدکلمه‌ها تجزیه می‌شود.

مثال ۶.۱ در مثال ۴.۱، کدگذاری دوتایی تاس یکتا دکدپذیر نیست. به‌عنوان نمونه $\mathbf{t} = ۱۱$ قابل تجزیه به فرم‌های ۱.۱ و ۱۱ (معرف $s = ۱۱$ و $s = ۳$) می‌باشد. این مشکل را با نمایش دوتایی ۳ رقمی سمبل‌ها می‌توان برطرف کرد:

$$۱ \longmapsto ۰۰۱, ۲ \longmapsto ۰۱۰, \ldots, ۶ \longmapsto ۱۱۰.$$

در این‌صورت $s = ۱۱ \longmapsto \mathbf{t} = ۰۰۱۰۰۱$، حال آن‌که $s = ۳ \longmapsto \mathbf{t} = ۰۱۱$. در حالت کلی قضیه زیر برقرار است.

قضیه ۷.۱ اگر تمامی کدکلمه‌های $w_i \in \mathcal{C}$ دارای طول یکسان باشند آنگاه $\mathcal{C}$ یکتا دکدپذیر است.

اثبات فرض کنید l طول مشترک کدکلمه‌ها باشد. اگر $\mathbf{t} \in \mathcal{C}^*$ دارای تجزیه $u_1 u_2 \ldots u_m = v_1 v_2 \ldots v_n$ باشد که $u_i, v_j \in \mathcal{C}$، آنگاه $lm = |t| = ln$، و در نتیجه $m = n$. چون u_1 و v_1 هر دو شامل l سمبل اول در $\mathbf{t}$ هستند داریم $u_1 = v_1$، و مشابهاً برای هر i تساوی $u_i = v_i$ برقرار است. ■

اگر تمامی کدکلمه‌ها طول مشترک l داشته باشند آنگاه $\mathcal{C}$ را یک کد قالبی به‌طول l می‌نامیم. این کدها به‌طور مفصل در فصل‌های ۵–۷ مطالعه می‌شوند. عکس قضیه ۷.۱ برقرار نیست.

مثال ۸.۱ کد دوتایی $\mathcal{C}$ با ضابطه

$$s_1 \longmapsto w_1 = ۰, \; s_2 \longmapsto w_2 = ۰۱, \; s_3 \longmapsto w_3 = ۰۱۱,$$

دارای طول متغیر بوده ولی یکتا دکدپذیر می‌باشد. در هر دنباله $\mathbf{t}$ یک ۰ معرف شروع یک کدکلمه w_i بوده که i یک واحد بیشتر از تعداد ۱ های بعدی است. برای نمونه اگر $\mathbf{t} = ۰۰۱۰۱۱۰۱۰۰۱۱$ آنگاه $\mathbf{s} = s_1 s_2 s_3 s_2 s_1 s_3$. در عمل ما از نماد $۰ \in T$ به‌عنوان یک علامت نقطه‌گذاری استفاده می‌کنیم.

در پی ارائه یک شرط لازم و کافی برای یکتا دکدپذیر بودن یک کد $\mathcal{C}$ هستیم. با استفاده از استقرا یک دنباله $\mathcal{C}_۰, \mathcal{C}_۱, \mathcal{C}_۲, \ldots$ از مجموعه‌های غیر خالی از کلمات در T^+ تعریف می‌کنیم. با قرار دادن $\mathcal{C}_۰ = \mathcal{C}$، برای $n \geq ۱$ تعریف می‌کنیم:

$$\mathcal{C}_n = \{w \in T^+ \mid v \in \mathcal{C}, u \in \mathcal{C}_{n-1} \text{ یا } v \in \mathcal{C}_{n-1}, u \in \mathcal{C} \text{ که } uw = v\}; \qquad (۳.۱)$$

سرانجام قرار می‌دهیم

$$\mathcal{C}_\infty = \bigcup_{n=۱}^{\infty} \mathcal{C}_n. \qquad (۴.۱)$$

این تعریف ممکن است در نگاه اول قدری سخت به‌نظر برسد، ولی با بیان مرحله‌به‌مرحله واضح‌تر خواهد شد. تعریف را با $\mathcal{C}_0 = \mathcal{C}$ شروع کرده و سپس $\mathcal{C}_n$ $(n \geq 1)$ را برحسب جمله قبلی آن یعنی $\mathcal{C}_{n-1}$ ساخته و سرانجام قرار می‌دهیم $\mathcal{C}_\infty = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots$. توجه کنید که برای $n = 1$ تعریف $\mathcal{C}_n$ را می‌توان ساده نمود؛ چون $\mathcal{C}_{n-1} = \mathcal{C}_0 = \mathcal{C}$ دو شرط تفکیک شده با کلمه ' یا' در (۳.۱) یکسان هستند. پس

$$\mathcal{C}_1 = \{w \in T^+ \mid uw = v \text{ و } u, v \in \mathcal{C} \text{ که }\}.$$

نیز توجه کنید که اگر $\mathcal{C}_{n-1} = \emptyset$ آنگاه $\mathcal{C}_n = \emptyset$ و بنابراین $\mathcal{C}_{n+1} = \mathcal{C}_{n+2} = \cdots = \emptyset$.

مثال ۹.۱ همچون مثال ۸.۱ فرض کنید $\mathcal{C} = \{0, 01, 011\}$. از (۳.۱) نتیجه می‌شود $\mathcal{C}_1 = \{1, 11\}$، زیرا از $011 = 01.1$ با $\mathcal{C} = \mathcal{C}_0 \in 01, 011$ نتیجه می‌شود $1 \in \mathcal{C}_1$ و از $011 = 0.11$ با خاصیت $\mathcal{C} = \mathcal{C}_0 \in 0, 011$ نتیجه می‌شود $11 \in \mathcal{C}_1$. در مرحله بعد، برای $n = 2$ ملاحظه می‌شود که کلمه‌ای چون $w \in T^+$ با خاصیت $uw = v$ که $u \in \mathcal{C}$ و $v \in \mathcal{C}_1$ وجود ندارد. پس $\mathcal{C}_2 = \emptyset$، و بنابراین $\mathcal{C}_n = \emptyset$ برای $n \geq 2$. پس از (۴.۱) نتیجه می‌شود $\mathcal{C}_\infty = \mathcal{C}_1 = \{1, 11\}$.

از تعریف $\mathcal{C}_\infty$ تصور می‌شود که ساخت این مجموعه به تعداد نامتناهی مرحله نیاز دارد. تمرین ۱.۱ نشان می‌دهد که همچون مثال ۹.۱ ساخت $\mathcal{C}_\infty$ در تعداد متناهی مرحله صورت می‌گیرد.

تمرین ۱.۱ ثابت کنید که اگر $\mathcal{C}$ کدکلمه‌های به‌طول $l_1, l_2, \ldots, l_q$ داشته و برای عددی چون n داشته باشیم $w \in \mathcal{C}_n$، آنگاه $|w| \leq l = \max\{l_1, \ldots, l_q\}$. نتیجه بگیرید که هر $\mathcal{C}_n$ متناهی بوده و دنباله $\mathcal{C}_0, \mathcal{C}_1, \ldots$ نهایتاً تناوبی می‌باشد. این چه کمکی در ساخت $\mathcal{C}_\infty$ ایجاد می‌کند؟

تمرین ۲.۱ مجموعه‌های $\mathcal{C}_n$ و $\mathcal{C}_\infty$ را برای کد سه‌تایی $\mathcal{C} = \{02, 12, 120, 20, 21\}$ بسازید. همین کار را برای $\mathcal{C} = \{02, 12, 120, 21\}$ انجام دهید.

اکنون یک شرط لازم و کافی برای یکتا دکدپذیری ارائه می‌دهیم. قضیه سردناس — پترسن[1] به شرح زیر است.

قضیه ۱۰.۱ کد $\mathcal{C}$ یکتا دکدپذیر است اگر و فقط اگر $\mathcal{C}$ و $\mathcal{C}_\infty$ فصل مشترک تهی داشته باشند.

قبل از درنظر گرفتن اثباتی برای قضیه، اجازه دهید این نتیجه را برای بعضی حالت‌های ساده

[1] Sardinas - Patterson theorem

به‌کار ببریم.

مثال ۱۱.۱ اگر همچون مثال ۸.۱ داشته باشیم $C = \{0, 01, 011\}$، آنگاه $C_\infty = \{1, 11\}$ که متمایز از C است. بنابراین از قضیه ۱۰.۱ یکتا دکدپذیری کد نتیجه می‌شود، همچنان‌که قبلاً ملاحظه شده است.

مثال ۱۲.۱ فرض کنید C کد سه‌تایی $\{0, 1, 2, 21, 210\}$ باشد. با به‌کار بردن (۳.۱) نتیجه می‌گیریم $C_1 = \{10\}$، $C_2 = \{0\}$ و $C_3 = \{1\}$، و بنابراین $1 \in C \cap C_\infty$. پس C یکتا دکدپذیر نیست (نیازی به محاسبه C_n برای $n > 3$ نیست). برای پیدا کردن مثالی از یکتا دکدناپذیری، توجه کنید که $10 \in C_1$ چون $2 \in C$ و $210 = 2.10 \in C$، سپس $0 \in C_2$ چون $1 \in C$ و $10 = 1.0 \in C_1$، و از آن‌جا $1 \in C_3$ چون $0 \in C_2$ و $01 = 0.1 \in C$. با کنار هم گذاشتن این تساوی‌ها نتیجه می‌گیریم

$$210.1 = 2.10.1 = 2.1.0.1 = 2.1.01,$$

و از این‌رو دنباله $\mathbf{t} = 2101$ می‌تواند به‌صورت‌های 210.1 و $2.1.01$ دکد شود.

تمرین ۳.۱ برای $C = \{02, 12, 120, 20, 21\}$ و $C = \{02, 12, 120, 21\}$ خاصیت یکتا دکدپذیری را بررسی کنید. اگر C یکتا دکدپذیر نیست یک کددنباله ارائه دهید که حداقل به دو دنباله متمایز دکد شود.

چون اثبات قضیه ۱۰.۱ نسبتاً طولانی است آن‌را در ضمیمه A ارائه کرده و در این‌جا به دو بحث تشریح کننده ایده‌های درگیر در قضیه می‌پردازیم.

($\Longrightarrow$) فرض کنید که $C \cap C_\infty \neq \emptyset$، و به‌عنوان نمونه فرض کنید $w \in C \cap C_2$. پس $uw = v$ که $u \in C$ و $v \in C_1$ و یا برعکس. برای سادگی فرض کنید حالت اول برقرار باشد (تمرین ۴.۱ را برای حالت دوم ملاحظه کنید). آنگاه $u'v = v'$ که $u', v' \in C$، و بنابراین دنباله $\mathbf{t} = u'uw \in T^*$ می‌تواند معرف یک دنباله $\mathbf{s}$ از سه سمبل منبع (چون $u', u, w \in C$) یا یک سمبل منبع (چون $u'uw = u'v = v' \in C$) باشد. پس عمل کدگشایی منحصربه‌فرد نیست.

($\Longleftarrow$) فرض کنید که نمونه‌ای از یکتا دکدناپذیری به‌فرم $\mathbf{t} = u_1u_2 = v_1v_2$ داشته باشیم که $u_1, u_2, v_1, v_2 \in C$. تساوی $|u_1| = |v_1|$ برقرار نیست زیرا از آن $u_1 = v_1$ و بنابراین $u_2 = v_2$ نتیجه می‌شود. بنابراین می‌توان فرض کرد $|u_1| > |v_1|$، و در نتیجه $u_1 = v_1w$ که $|w| > 0$. پس $w \in C_1$ و بنابراین $u_2 \in C_2$ زیرا $wu_2 = v_2$. از این‌رو $u_2 \in C \cap C_\infty$.

تمرین ۴.۱ فرض کنید $w \in C \cap C_2$ که $uw = v$ و $v \in C$، $u \in C_1$. یک کددنباله ارائه دهید که به بیش از یک دنباله دکد شود.

تمرین ۵.۱ کد $\mathcal{C}$ دارای یک کددنباله قابل تجزیه به فرم‌های ۰۱.۲۱۲.۰۱.۲۰ و ۰۱۲۱۲۰.۱۲۰ است. عضوی از $\mathcal{C} \cap \mathcal{C}_\infty$ را پیدا کنید.

تمرین ٦.١ فرض کنید که $w \in \mathcal{C} \cap \mathcal{C}_3$. با درنظر گرفتن حالت‌های مختلف ممکن برای $w \in \mathcal{C} \cap \mathcal{C}_3$ کددنباله‌های مختلفی ارائه دهید که کدگشایی یکتا نداشته باشند.

ایده‌های اصلی در اثبات قضیه ۱۰.۱ مشابه بحث‌هایی هستند که به آن‌ها اشاره شد ولی به‌جهت این‌که باید بینهایت حالت مختلف درنظر گرفته شود از پیچیدگی بیشتری برخوردار هستند. خوشبختانه یک شرط لازم و کافی ساده‌تری برای گروهی از کدها وجود دارد که در بخش بعد بررسی می‌شوند.

برای این‌که کددنباله‌های متناهی به‌صورت یکتا دکد شوند از مفهوم یکتا دکدپذیری استفاده کردیم. برای تعمیم این خاصیت به تمامی کددنباله‌ها، متناهی و غیرمتناهی، به الزام قوی‌تری نیاز است.

قضیه‌ای منتسب به ایون[۱] [Ev 63]، لونشتاین[۲] [Le 64] و رایلی[۳] [Ri 67] نشان می‌دهد که این خاصیت برقرار است اگر و فقط اگر $\mathcal{C} \cap \mathcal{C}_\infty = \emptyset$ و به‌ازای عددی چون $n \geq 1$ داشته باشیم $\mathcal{C}_n = \emptyset$. (این‌ها همچنین شروط لازم و کافی برای این هستند که $\mathcal{C}$ یکتا دکدپذیر با تأخیر محدود باشد، یعنی این‌که عددی ثابت چون d وجود دارد به‌قسمی‌که اگر دو کددنباله در d سمبل اول خود یکسان باشند آنگاه اولین کدکلمه آن‌ها مشترک می‌باشد؛ از این‌رو عمل کدگشایی بعد از حداکثر d سمبل شروع می‌شود. در بخش بعد یک شرط قوی‌تری را درنظر خواهیم گرفت.)

مثال ۱۳.۱ در مثال ۹.۱، هر دو شرط تأمین شده‌اند، و از این‌رو هر کددنباله‌ای به‌صورت یکتا دکد می‌شود. برای دیدن مثالی که در آن تمامی کددنباله‌های متناهی یکتا دکدپذیر بوده ولی بعضی از کددنباله‌های نامتناهی واجد این خاصیت نباشند به تمرین ۷.۱ مراجعه کنید.

تمرین ۷.۱ برای هر یک از کدهای سه‌تایی $\mathcal{C}$ در تمرین ۲.۱، مشخص کنید که آیا تمامی کددنباله‌های نامتناهی یکتا دکدپذیر بوده، و در صورتی که این خاصیت را ندارند مثالی از کددنباله‌های یکتا دکدناپذیر ارائه دهید.

در باقیمانده مطالب این کتاب توجه خود را محدود به کددنباله‌های متناهی می‌کنیم.

۳.۱ کدهای لحظه‌ای

قبل از تعریف کدهای لحظه‌ای به بررسی چند مثال می‌پردازیم.

---

[۱] Even [۲] Levenshtein [۳] Riley

مثال ۱۴.۱ کد دوتایی $\mathcal{C}$ را درنظر بگیرید که $s_1 \longmapsto ۰,\ s_2 \longmapsto ۰۱,\ s_3 \longmapsto ۱۱$. با به‌کار بردن نمادهای بخش ۲.۱ داریم $\mathcal{C}_1 = \mathcal{C}_2 = \cdots = \{۱\}$، و از این‌رو $\mathcal{C}_\infty = \{۱\}$؛ پس $\mathcal{C} \cap \mathcal{C}_\infty = \emptyset$ و در نتیجه بنابر قضیه ۱۰.۱ کد $\mathcal{C}$ یکتا دکدپذیر است. فرض کنید که یک دنباله متناهی باشروع ۰۱۱۱ دریافت شده است یعنی $\mathbf{t} = ۰۱۱۱\ldots$. اگرچه این کد یکتا دکدپذیر است ولی عمل کدگشایی تا قبل از رسیدن به انتهای قالب ۱ها انجام نمی‌شود. اگر تعداد ۱های این قالب زوج باشد t دارای تجزیه $\mathbf{t} = ۰.۱۱.۱۱\ldots$ بوده و پیام دکدشده به‌فرم $\mathbf{s} = s_1 s_3 s_3 \ldots$ می‌باشد؛ ولی اگر تعداد ۱ ها فرد باشد آنگاه t دارای تجزیه $\mathbf{t} = ۰۱.۱۱.۱۱\ldots$ و پیام دکدشده به‌شکل $\mathbf{s} = s_2 s_3 s_3 \ldots$ می‌باشد. در شرایط خاص، این تأخیر در کدگشایی مشکلاتی را ایجاد می‌کند. گفته می‌شود که $\mathcal{C}$ یک کد لحظه‌ای نیست.

مثال ۱۵.۱ متن ... RUSSIANS DECLARE WAR توسط تلکس نخست‌وزیر چاپ می‌شود. یک تصمیم سریع گرفته می‌شود، یک دکمه فشار داده می‌شود، و در عرض چند دقیقه انفجارهای بزرگی صورت می‌گیرد. خیلی زود هر کسی کشته شده است. در این مدت تلکس کار چاپ را ادامه می‌دهد

... RINGTON VODKA TO BE EXCELLENT.

تمرین ۸.۱ نشان دهید که کد دوتایی $\mathcal{C} = \{۰, ۰۱, ۰۱۱, ۱۱۱\}$ یکتا دکدپذیر است؛ واکنش یک دریافت کننده نسبت به دنباله دریافتی $۰۱۱۱\ldots۱\ldots$ چگونه باید باشد؟

مثال ۱۶.۱ کد دوتایی $\mathcal{D}$ را درنظر بگیرید که $s_1 \longmapsto ۰,\ s_2 \longmapsto ۱۰,\ s_3 \longmapsto ۱۱$؛ این کد معکوس کد تمرین ۱۴.۱ است. یکتا دکدپذیری این کد را می‌توان با قضیه ۱۰.۱ و یا از کد $\mathcal{C}$ نتیجه گرفت. این کد لحظه‌ای نیز می‌باشد، به این معنی که می‌توان دنباله t را با حرکت به‌جلو دکد کرد: یک ۰ به معنی w_1 و بنابراین s_1 بوده، و یک ۱ اشاره به شروع $w_2 = ۱۰$ یا $w_3 = ۱۱$ دارد، که در این حالت به محض ملاحظه سمبل بعدی به s_2 یا s_3 دکد می‌شود. بنابراین در دنباله t هر کدکلمه‌ای به‌محض رویت شدن و بدون هیچ تأخیری دکد می‌شود.

تمرین ۹.۱ آیا این خاصیت برای کد $\mathcal{D} = \{۰, ۱۰, ۱۱۰, ۱۱۱\}$ نیز که معکوس کد $\mathcal{C}$ معرفی شده در مثال ۸.۱ است برقرار است؟

به‌صورت رسمی کد $\mathcal{C}$ لحظه‌ای نامیده می‌شود اگر برای هر دنباله از کدکلمه‌های $w_{i_1}, w_{i_2}, \ldots, w_{i_n}$ هر کددنباله به‌فرم $\mathbf{t} = w_{i_1} w_{i_2} w_{i_3} \ldots w_{i_n} \ldots$ به‌طور یکتا به $\mathbf{s} = s_{i_1} s_{i_2} \ldots s_{i_n} \ldots$ دکد شود صرف‌نظر از این‌که سمبل‌های بعدی در t چه هستند. بنابراین کد $\mathcal{C}$ در مثال ۱۴.۱ لحظه‌ای نیست؛ یک دنباله $\mathbf{t} = w_1 w_3 \ldots = ۰۱۱\ldots$ بسته به سمبل‌های بعدی ممکن است به $\mathbf{s} = s_1 s_3 \ldots$ و یا $\mathbf{s} = s_2 s_3$ دکد شود. کد D در مثال

۱۶.۱ لحظه‌ای است؛ به‌محض دریافت $w_{i_1}w_{i_2}\dots w_{i_n}$ می‌دانیم که این قسمت معرف $s_{i_1}s_{i_2}\dots s_{i_n}$ است و این بستگی به سمبل‌های بعدی ندارد. بنابه تعریف، هر کد لحظه‌ای یکتا دکدپذیر است؛ مثال ۱۴.۱ نشان می‌دهد که عکس این حکم نادرست است.

یک کد $\mathcal{C}$ را کد پیشوند[1] نامیم هرگاه هیچ کدکلمه‌ای چون w_i پیشوند کدکلمه دیگری چون w_j نباشد؛ معادلاً $w_j \neq w_i w$ که $w \in T^*$، یعنی این‌که با نمادهای بخش ۲.۱ داریم $\mathcal{C}_1 = \emptyset$. پس در مثال ۱۴.۱ کد $\mathcal{C}$ پیشوند نیست (زیرا ۰ پیشوند ۰۱ است)، ولی کد معکوس شده D در مثال ۱۶.۱ یک کد پیشوند است.

قضیه ۱۷.۱ یک کد $\mathcal{C}$ لحظه‌ای است اگر و فقط اگر پیشوند باشد.

اثبات ($\Longrightarrow$) اگر $\mathcal{C}$ پیشوند نباشد، مثلاً w_i یک پیشوند w_j باشد، آنگاه کددنباله $\mathrm{t} = w_i \dots$ ممکن است به $\mathrm{s} = s_i \dots$ یا $\mathrm{s} = s_j \dots$ دکد شود، و بنابراین $\mathcal{C}$ لحظه‌ای نیست.
($\Longleftarrow$) اگر $\mathcal{C}$ یک کد پیشوند باشد و t با $w_i \dots$ شروع شود، آنگاه s باید با s_i شروع شود زیرا هیچ کدکلمه $w_j (j \neq i)$ وجود ندارد که پیشوند w_i باشد و یا w_i پیشوند w_j. با ادامه کار به همین روش هر کدکلمه دریافتی در t را دکد می‌کنیم. بنابراین $\mathcal{C}$ لحظه‌ای است.■

مثال‌های ۱۴.۱ و ۱۶.۱ این نتیجه را توضیح می‌دهند.

۴.۱ ساخت کدهای لحظه‌ای

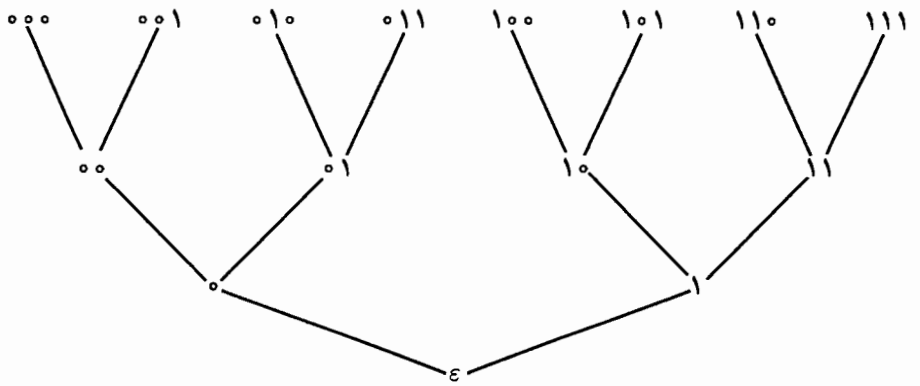
به‌منظور ساخت کدهای لحظه‌ای درنظر گرفتن مجموعه T^* حاوی کلمات در T به‌صورت یک گراف مفید خواهد بود. یک گراف مجموعه‌ای از نقاط (به نام رئوس) است که بعضی از آن‌ها توسط اضلاع به‌هم وصل هستند. در این حالت، رأس‌ها همان کلمات $w \in T^*$ بوده و هر w با یک ضلع به کلمات $wt_1, wt_2, \dots, wt_r$ متصل است که به‌وسیله اضافه شدن $t_i \in T$ به انتهای w ساخته می‌شود. این گراف را می‌توان با رشد به سمت بالا و با ریشه ε، کلمه خالی، درنظر گرفت که در آن کلمات به‌طول l در لایه l ام بالای ε قرار دارند. در نظریه گراف، این گراف یک درخت r–تایی ریشه‌دار نامیده می‌شود.(یک درخت گرافی همبند است که حاوی دور نباشد؛ در این‌جا ε ریشه است.)

شکل ۱.۱ درخت دوتایی T^* را تا لایه سوم با $T = \mathbf{Z}_2$ نمایش می‌دهد.

تمرین ۱۰.۱ درخت سه‌تایی T^* را تا لایه سوم، $l = 3$، با $T = \mathbf{Z}_3$ رسم کنید.

یک کد $\mathcal{C}$ را می‌توان به‌عنوان یک مجموعه متناهی از رئوس درخت T^* درنظر گرفت. یک کلمه w_i پیشوند کلمه w_j است اگر و فقط اگر رأس w_i مغلوب رأس w_j باشد، یعنی این‌که یک مسیر به سمت بالا در T^* از w_i به w_j وجود داشته باشد. از این‌رو از قضیه ۱۷.۱

[1] prefix

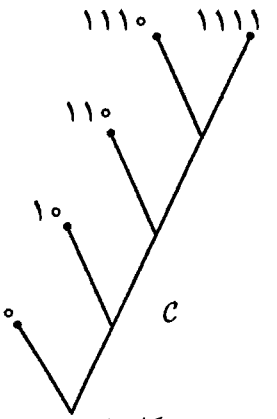


شکل ۱.۱

نتیجه می‌شود که $\mathcal{C}$ لحظه‌ای است اگر و فقط اگر هیچ رأس $w_i \in \mathcal{C}$ مغلوب رأس دیگری چون $w_j \in \mathcal{C}(i \neq j)$ نباشد. این محک را می‌توان برای ساخت کدهای لحظه‌ای به‌کار برد؛ این عمل با انتخاب رئوس T^* به‌قسمی‌که رأس انتخابی فعلی غالب بر رئوس قبلاً انتخاب شده نباشد صورت می‌گیرد.

مثال ۱۸.۱ می‌خواهیم یک کد لحظه‌ای $\mathcal{C}$ برای منبع $\mathcal{S}$ با پنج سمبل s_1،...،s_5 بسازیم. ابتدا $s_1 \longmapsto w_1 = 0$ را امتحان می‌کنیم، پس ۰ یک رأس $\mathcal{C}$ است. اگر قرار است که $\mathcal{C}$ پیشوند باشد آنگاه هیچ رأس دیگری از $\mathcal{C}$ نمی‌تواند بر ۰ غالب باشد، بنابراین آن‌ها باید غالب بر ۱ باشند (یعنی این‌که آن‌ها باید با ۱ شروع شوند). اگر $s_2 \longmapsto w_2 = 1$ را انتخاب کنیم آنگاه کدکلمه بیشتری نمی‌توان به کد اضافه کرد زیرا w_1 یا w_2 را مغلوب خواهند کرد. در عوض اجازه دهید $s_2 \longmapsto w_2 = 10$ را امتحان کنیم. در این‌صورت $s_3 \longmapsto w_3 = 11$ غیرممکن است زیرا اجازه وجود کدکلمه بیشتری را نمی‌دهد، از این‌رو انتخاب $s_3 \longmapsto w_3 = 110$ را بررسی می‌کنیم. با این روش به انتخاب‌های $s_4 \longmapsto w_4 = 1110$ و $s_5 \longmapsto w_5 = 1111$ می‌رسیم. این یک کد لحظه‌ای دوتایی $\mathcal{C} = \{0, 10, 110, 1110, 1111\}$ با طول کدهای $l_i = 1, 2, 3, 4, 4$ ایجاد می‌کند که در شکل ۲.۱ نشان داده شده است. (این تنها حالت ممکن نیست؛ به‌عنوان نمونه کد دوتایی $\{00, 01, 10, 110, 111\}$ نیز لحظه‌ای است.)

مثال ۱۹.۱ آیا برای این منبع یک کد دوتایی لحظه‌ای با طول کدهای ۱، ۲، ۳، ۳، ۴ وجود دارد؟ مجدداً درخت دوتایی T^* را به‌کار می‌بریم. انتخاب یک کدکلمه w_1 به‌طول $l_1 = 1$، یعنی یک رأس به ارتفاع ۱، نصف رئوس T^* را از انتخاب‌های بعدی حذف می‌کند؛ در واقع رئوس غالب بر w_1 حذف می‌شوند، در نتیجه نصف رئوس باقی می‌مانند. یک انتخاب به ارتفاع ۲ برای w_2 نیمی دیگر از رئوس باقیمانده، $\frac{1}{4}$، را حذف می‌کند، و بنابراین $1 - \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$ رئوس T^* باقی می‌مانند. هر انتخابی برای w_3 و w_4 در لایه



شکل ۲.۱

سوم باعث حذف $\frac{۱}{۴} = \frac{۱}{۲۴} + \frac{۱}{۲۴}$ رئوس T^* شده و از این‌رو انتخابی برای $w_۵$ باقی نمی‌ماند. مشکل این است که مجموع درصد رئوسی از T^* که بالای کلمات w_i قرار می‌گیرند از ۱ تجاوز می‌کند؛

$$\frac{۱}{۲} + \frac{۱}{۲۲} + \frac{۱}{۲۳} + \frac{۱}{۲۳} + \frac{۱}{۲۴} > ۱.$$

بنابراین هیچ کد لحظه‌ای دوتایی با طول کدهای ۱، ۲، ۳، ۳، ۴ برای منبع S وجود ندارد. با این حال یک کد لحظه‌ای سه‌تایی با این طول‌ها وجود دارد: اگر $r = ۳$ آنگاه انتخاب‌های $w_۱، w_۲، \ldots، w_۵$ باعث حذف نسبت‌های $\frac{۱}{۳}$، $\frac{۱}{۹}$، $\frac{۱}{۲۷}$، $\frac{۱}{۲۷}$، $\frac{۱}{۸۱}$ از رئوس درخت سه‌تایی T^* می‌شوند که $|T| = ۳$. چون

$$\frac{۱}{۳} + \frac{۱}{۹} + \frac{۱}{۲۷} + \frac{۱}{۲۷} + \frac{۱}{۸۱} \leq ۱,$$

چنین انتخاب‌هایی امکان‌پذیر هستند.

تمرین ۱۱.۱ یک کد لحظه‌ای سه‌تایی با طول کدهای ۱، ۲، ۳، ۳، ۴ پیدا کنید. آیا چنین کدی با طول‌های ۱، ۱، ۲، ۲، ۲ وجود دارد؟

این مفهوم 'نسبت' از یک درخت نامتناهی T^* مفید بوده ولی دقیق نیست. با دقیق‌تر کردن آن، می‌توانیم مانند قبل شرایط لازم و کافی برای وجود کدهای لحظه‌ای r–تایی با طول کدهای داده شده ارائه دهیم.

۵.۱ نامساوی کرفت

متأثر از مثال‌های بخش ۴.۱، نتیجه زیر را داریم که به نامساوی کرفت[۱] [Kr 49] شناخته شده است.

[۱] Kraft's inequality

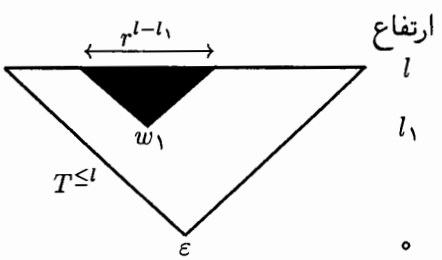
قضیه ۲۰.۱ یک کد لحظه‌ای r–تایی $\mathcal{C}$ با طول کدهای $l_1, l_2, \ldots, l_q$ وجود دارد اگر و فقط اگر

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} \leq 1. \qquad (5.1)$$

اثبات ($\Longleftarrow$) با شماره‌گذاری مجدد طول کدها (در صورت نیاز) به خاصیت $l_1 \leq l_2 \leq \cdots \leq l_q$ می‌رسیم. فرض کنید $l = \max(l_1, \ldots, l_q)$ و بخشی از T^* که تا ارتفاع l قرار دارد را درنظر بگیرید: $T^{\leq l} = T^0 \cup T^1 \cup \cdots \cup T^l$. این یک درخت متناهی است: این درخت در هر ارتفاع $h = 0, 1, \ldots, l$ دارای r^h رأس است که اعضای T^h، یعنی کلمات به‌طول h، می‌باشند. برگ‌های آن شامل r^l رأس در ارتفاع l هستند.

یک رأس w_1 دلخواه در ارتفاع l_1 را به s_1 اختصاص داده و یک کدکلمه به‌طول l_1 به‌دست آورده و سپس w_1 و تمامی قسمت‌هایی از $T^{\leq l}$ که در بالای w_1 قرار دارند را حذف می‌کنیم (زیرا آن رئوس دیگر قابل استفاده نخواهد بود)؛ این عمل بالاخص باعث حذف r^{l-l_1} برگ شده که معرف کلمات به‌طول $l - l_1$ با شروع w_1 می‌باشند (شکل ۳.۱ را ببینید). اگر $q > 1$ آنگاه بنابر (۵.۱) داریم:

$$r^{l-l_1} < r^l \sum_{i=1}^{q} \frac{1}{r^{l_i}} \leq r^l.$$



شکل ۳.۱

بنابراین حداقل یک برگ از $T^{\leq l}$ حذف نشده است. اولین l_2 سمبل این برگ یک کدکلمه w_2 به‌طول l_2 ایجاد می‌کند که بالای w_1 قرار ندارد. اکنون w_2 و تمامی قسمت‌های $T^{\leq l}$ که بالای آن قرار دارند را حذف می‌کنیم و بنابراین r^{l-l_2} برگ دیگر حذف می‌شوند (هیچ برگی نمی‌تواند هم بالای w_1 و هم بالای w_2 باشد). با ادامه این روش، انتخاب کدکلمه و حذف‌کردن، هیچ کدکلمه‌ای بالای یا زیر کدکلمه دیگری قرار نخواهد داشت. بعد از این‌که k کدکلمه w_1، ...، w_k انتخاب شده‌اند، $k < q$، تعداد برگ‌های حذف‌شده برابر

$r^{l-l_1}+\cdots+r^{l-l_k}$ است. چون

$$\sum_{i=1}^{k} r^{l-l_i} < r^l \sum_{i=1}^{q} \frac{1}{r^{l_i}} \leq r^l,$$

پس حداقل یک برگ باقی مانده و می‌توان کدکلمه w_{k+1} به‌طول l_{k+1} را به‌عنوان پیشوند این برگ انتخاب کرد. بنابراین می‌توانیم انتخاب کدکلمه‌ها را تا تعیین w_q ادامه دهیم. در تمامی این مراحل شرط پیشوندی برقرار بوده و از این‌رو بنابر قضیه ۱۷.۱ کد به‌دست آمده $\mathcal{C}=\{w_1,\ldots,w_q\}$ لحظه‌ای است.

($\Longrightarrow$) اگر $\mathcal{C}$ لحظه‌ای باشد آنگاه بنابر قضیه ۱۷.۱ یک کد پیشوند است، از این‌رو هر برگ از $T^{\leq l}$ بالای حداکثر یک کدکلمه از $\mathcal{C}$ قرار دارد. هر کدکلمه w_i از $\mathcal{C}$ پایین r^{l-l_i} برگ قرار دارد (که $l_i=|w_i|$)، و بنابراین با محاسبه مجموع روی تمامی کدکلمه‌های $w_i\in\mathcal{C}$ ملاحظه می‌کنیم که تعداد برگ‌های بالای کدکلمه‌ها برابر $\sum_i r^{l-l_i}$ است. چون تعداد برگ‌های $T^{\leq l}$ برابر r^l است باید داشته باشیم

$$\sum_{i=1}^{q} r^{l-l_i} \leq r^l.$$

از تقسیم طرفین این رابطه بر r^l نامساوی (۵.۱) نتیجه می‌شود. ■

برای مثال‌های توضیح دهنده نامساوی کرفت به بخش ۴.۱ مراجعه کنید.

۶.۱ نامساوی مک‌میلان

ملاحظه کرده‌ایم که کلاس کدهای یکتا دکدپذیر اکیداً بزرگ‌تر از کلاس کدهای لحظه‌ای است، از این‌رو ممکن است این توقع ایجاد شود که یک شرط لازم و کافی ضعیف‌تری در مقایسه با (۵.۱) برای وجود کدهای r–تایی یکتا دکدپذیر با طول کدهای داده شده وجود داشته باشد. واقعیت شگفت‌انگیز این است که چنین حکمی برقرار نیست. نامساوی مک‌میلان[1] به شرح زیر است.

قضیه ۲۱.۱ یک کد r–تایی یکتا دکدپذیر $\mathcal{C}$ با طول کدهای $l_1,\ldots,l_q$ وجود دارد اگر و فقط اگر

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} \leq 1. \qquad (6.1)$$

اثبات ($\Longleftarrow$) بنابر قضیه ۲۰.۱ یک کد لحظه‌ای r–تایی با طول کدهای داده شده وجود دارد، و چون کد لحظه‌ای است پس یکتا دکدپذیر است.

($\Longrightarrow$) فرض کنید $\mathcal{C}$ کد r–تایی یکتا دکدپذیر با طول کدهای $l_1,\ldots,l_q$ باشد. تعریف کنید

[1] Mc Millan's inequality

$$K = \sum_{i=1}^{q} \frac{1}{r^{l_i}}, \tag{۷.۱}$$

و قرار دهید:

$$l = \max(l_1, \ldots, l_q),$$
$$m = \min(l_1, \ldots, l_q).$$

اکنون بسط K^n را درنظر بگیرید که $n \geq 1$:

$$K^n = \left(\sum_{i=1}^{q} \frac{1}{r^{l_i}} \right)^n.$$

این بسط برابر مجموع جملاتی به‌فرم

$$\frac{1}{r^{l_{i_1}}} \times \frac{1}{r^{l_{i_2}}} \times \cdots \times \frac{1}{r^{l_{i_n}}} = \frac{1}{r^j}$$

است که در آن

$$j = l_{i_1} + \cdots + l_{i_n}. \tag{۸.۱}$$

چون برای هر i نامساوی $m \leq l_i \leq l$ برقرار است پس $mn \leq j \leq ln$، و بنابراین

$$K^n = \sum_{j=mn}^{ln} \frac{N_{j,n}}{r^j}.$$

در این عبارت عدد $N_{j,n}$ (ضریب $\frac{1}{r^j}$) تعداد حالت‌هایی است که می‌توان j را به‌فرم (۸.۱) به‌صورت مجموع n طول کلمه (احتمالاً با تکرار) بیان کرد؛ معادلاً $N_{j,n}$ معرف تعداد دنباله‌های w_{i_1}، ...، w_{i_n} از n کدکلمه از C است که مجموع طول آن‌ها برابر j است. هر یک چنین دنباله‌ای یک کددنباله $\mathbf{t} = w_{i_1} \ldots w_{i_n}$ به‌طول j مشخص می‌کند، و چون C یکتا دکدپذیر است هر $\mathbf{t}$ حداکثر از یک دنباله از این نوع کدکلمه‌ها منتج می‌شود. بنابراین $N_{j,n}$ حداکثر برابر تعداد کددنباله‌های $\mathbf{t}$ به‌طول j بوده و بنابراین $N_{j,n} \leq r^j$. چون K^n مجموع $ln + 1 - mn$ جمله $\frac{N_{j,n}}{r^j} \leq 1$ است برای هر $n \geq 1$ داریم:

$$K^n \leq (l - m)n + 1. \tag{۹.۱}$$

چون K، l، و m مستقل از n هستند، اگر $K > 1$ آنگاه سمت چپ این نامساوی رشد نمایی داشته درحالی‌که سمت راست به‌صورت خطی افزایش می‌یابد. این خاصیت برای n های به‌قدر کافی بزرگ در تناقض با (۹.۱) بوده و از این‌رو $K \leq 1$. ■

اثبات فوق متعلق به کاروش[۱] [Ka 61] است؛ در اثبات اولیه از توابع مختلط استفاده شده است (به [McM 56] یا [Re 61، صفحات 8-147] مراجعه کنید). حکم زیر نتیجه فوری قضایای ۲۰.۱ و ۲۱.۱ است.

نتیجه ۲۲.۱ یک کد r–تایی لحظه‌ای با طول کدهای $l_1,\ldots,l_q$ وجود دارد اگر و فقط اگر یک کد یکتا دکدپذیر r–تایی با این طول‌ها موجود باشد.

۷.۱ نکاتی راجع به نامساوی‌های کرفت و مک‌میلان

نکته ۲۳.۱ قضایای ۲۰.۱ و ۲۱.۱ نمی‌گویند که یک کد r–تایی با طول‌های $l_1,\ldots,l_q$ لحظه‌ای یا یکتا دکدپذیر است اگر و فقط اگر $\sum r^{-l_i} \leq ۱$. به‌عنوان نمونه کد دوتایی $\mathcal{C} = \{۰, ۰۱, ۰۱۱\}$ دارای طول‌های $l_i = ۱, ۲, ۳$ بوده و $\sum r^{-l_i} = \frac{۷}{۸} \leq ۱$؛ ولی $\mathcal{C}$ پیشوند نبوده و در نتیجه لحظه‌ای نیست. مشابهاً می‌توان یک کد دوتایی با این طول‌ها ارائه کرد که یکتا دکدپذیر نباشد؛ یک مثال واضح کد $\{۰, ۰۱, ۰۰۱\}$ است.

نکته ۲۴.۱ قضایای ۲۰.۱ و ۲۱.۱ اظهار می‌دارند که اگر $\sum r^{-l_i} \leq ۱$ آنگاه کدهایی با این طول‌ها وجود دارند که لحظه‌ای و در نتیجه یکتا دکدپذیر هستند. به‌عنوان نمونه کد دوتایی $\{۰, ۱۰, ۱۱۰\}$ یک کد پیشوند بوده و از این‌رو واجد هر دو شرط می‌باشد.

نکته ۲۵.۱ یک کد r–تایی یکتا دکدپذیر $\mathcal{C}$ لزوماً لحظه‌ای نیست، ولی بنابر نتیجه ۲۲.۱ باید یک کد r–تایی لحظه‌ای با این پارامترها وجود داشته باشد. به‌عنوان نمونه، کد دوتایی $\mathcal{C} = \{۰, ۰۱, ۱۱\}$ که در مثال ۱۴.۱ معرفی شد یکتا دکدپذیر است ولی لحظه‌ای نیست؛ با همین طول کدها ما کد لحظه‌ای $\mathcal{D} = \{۰, ۱۰, ۱۱\}$، مثال ۱۶.۱، را داریم.

نکته ۲۶.۱ عبارت r^{-l_i} در $K = \sum r^{-l_i}$ متناظر است با بخشی از درخت T^* که در بالای w_i، یک رأس به ارتفاع l_i قرار دارد. این تفسیر لزوم $K \leq ۱$ را توضیح می‌دهد.

۸.۱ تمرینات تکمیلی

تمرین ۱۲.۱ آیا یک کد لحظه‌ای سه‌تایی (روی $\mathbf{Z}_۳$) با طول کدهای ۱،۲،۲،۲،۲،۲،۲،۳،۳،۳ و ۳،۳،۳،۳،۲،۲،۲،۲،۲،۱ وجود دارد؟ یک کد با طول‌های $l_i = ۱, ۲, ۲, ۲, ۲, ۲, ۳, ۳, ۳$ بسازید؛ چند کد این‌چنینی وجود دارد؟

تمرین ۱۳.۱ از کد دوتایی $\mathcal{C} = \{۰, ۱۰, ۱۱\}$ استفاده می‌شود؛ چند کددنباله به‌طول j از

[۱] Karush

روی $\mathcal{C}$ می‌توان ساخت؟ (راهنمایی: یک رابطه بازگشتی برای این عدد N_j پیدا کرده و آن‌را حل کنید.)

تمرین ۱۴.۱ فرض کنید $|T| = r$، $l_1 \leq \cdots \leq l_q$، و $\sum r^{-l_i} \leq 1$. به چند طریق یک نفر می‌تواند کلمات $w_1, \ldots, w_q$ از T^* را انتخاب کند به‌قسمی‌که $|w_i| = l_i$ و $\{w_1, \ldots, w_q\}$ یک کد لحظه‌ای باشد؟

تمرین ۱۵.۱ یک کد را جامع می‌نامند هرگاه هر دنباله به‌قدر کافی طولانی از کدسمبل‌ها با یک کدکلمه شروع شود (معادلاً، هر دنباله نامتناهی از سمبل‌های کد قابل تجزیه به کدکلمه‌ها باشد). یک شرط معادل برحسب تعداد برگ‌های درخت $T^{\leq l}$ که در اثبات قضیه ۲۰.۱ به‌کار رفت ارائه دهید. کدام‌یک از کدهای مثال‌های این فصل جامع هستند؟

تمرین ۱۶.۱ نشان دهید اگر یک کد r–تایی $\mathcal{C}$ با طول‌های $l_1, \ldots, l_q$ جامع باشد آنگاه $\sum r^{-l_i} \geq 1$، و تساوی برقرار است اگر و فقط اگر $\mathcal{C}$ لحظه‌ای باشد.

تمرین ۱۷.۱ فرض کنید $\mathcal{C}$ یک کد r–تایی با طول‌های $l_1, \ldots, l_q$ باشد. نشان دهید هر دو خاصیت از خواص سه‌گانه زیر خاصیت سوم را ایجاب می‌کنند:

(الف) $\mathcal{C}$ لحظه‌ای است؛

(ب) $\mathcal{C}$ جامع است؛

(ج) $\sum r^{-l_i} = 1$.

نشان دهید که هیچ‌یک از این‌ها از یکی از دو خاصیت دیگر نتیجه نمی‌شود.

فصل ۲

کدهای بهینه

آدم‌های کم‌حرف بهترین انسان‌ها هستند.(King Henry V)

در فصل ۱ نحوه کدگذاری اطلاعات به‌قسمی‌که عمل کدگشایی یکتا و یا لحظه‌ای باشد را ملاحظه کردیم. در هر یک از این دو حالت نیاز اصلی، براساس نامساوی کرفت و یا مک‌میلان، این است که از کدکلمه‌های به‌قدر کافی طولانی استفاده شود. این نکته مسئله کارایی را پیش می‌آورد: اگر کدکلمه‌ها خیلی بزرگ باشند آنگاه ممکن است عمل ذخیره‌سازی دشوار بوده و عمل انتقال به‌کندی صورت گیرد. بنابراین ما باید یک تعادلی ایجاد کنیم به‌قسمی‌که کدکلمه‌ها به‌قدر کافی بزرگ بوده تا عمل کدگشایی سریع باشد و از طرف دیگر طول کدکلمه‌ها به‌قدر کافی کوتاه بوده تا از نقطه‌نظر اقتصادی مقرون به صرفه باشد. از این جهت بهترین کدهای موجود را کدهای بهینه نامیده که همان کدهای لحظه‌ای با کم‌ترین متوسط طول کد می‌باشند. نشان می‌دهیم که این کدها وجود داشته و از الگوریتم هافمن[۱] برای ساخت آن‌ها استفاده می‌شود. به‌جهت سادگی، عمدتاً روی کدهای دوتایی بحث کرده، اگرچه به‌طور خلاصه نحوه تعمیم این ایده‌ها را روی حالت غیر دوتایی توضیح خواهیم داد.

[۱] Huffman's algorithm

۱.۲ بهینگی

همچون فصل ۱ فرض کنید $\mathcal{S}$ یک منبع باشد. فرض می‌کنیم که احتمال‌های

$$p_i = Pr(X_n = s_i) = Pr(s_i)$$

مستقل از زمان n و سمبل‌های قبلی X_1، ...، X_{n-1} باشند. نظریه‌ای که ذیلاً ارائه می‌شود قابل تعمیم به منابعی است که این خواص را ندارند ولی ما روی ساده‌ترین حالتی متمرکز می‌شویم که در آن این شرایط برقرار هستند. چون اعداد p_i تشکیل یک توزیع احتمال می‌دهند داریم

$$0 \leq p_i \leq 1 \ , \quad \sum_{i=1}^{q} p_i = 1. \tag{۱.۲}$$

اگر کد $\mathcal{C}$ برای منبع $\mathcal{S}$ دارای طول کدهای l_1، ...، l_q باشد آنگاه میانگین طول کد برابر است با:

$$L = L(\mathcal{C}) = \sum_{i=1}^{q} p_i l_i. \tag{۲.۲}$$

واضح است که برای هر کد $\mathcal{C}$ نامساوی $L(\mathcal{C}) > 0$ برقرار است. به‌منظور اقتصادی بودن و کارایی، سعی می‌کنیم با حفظ خاصیت کدگشایی لحظه‌ای عدد $L(\mathcal{C})$ را تا حد ممکن کوچک کنیم. یعنی این‌که برای عدد داده شده r و توزیع احتمال (p_i)، سعی می‌کنیم یک کد r–تایی لحظه‌ای $\mathcal{C}$ پیدا کنیم که دارای کم‌ترین مقدار $L(\mathcal{C})$ باشد. چنین کدهایی بهینه یا فشرده[1] نامیده می‌شوند.

مثال ۱.۲ فرض کنید $\mathcal{S}$ هوای روزانه (مثال ۲.۱) باشد با $p_i = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ برای $i = 1, 2, 3$. کد دوتایی $\mathcal{C} = \{00, 01, 1\}$ لحظه‌ای است (زیرا یک کد پیشوند است) و

$$L(\mathcal{C}) = \frac{1}{4}.2 + \frac{1}{2}.2 + \frac{1}{4}.1 = 1.75;$$

ولی کد دوتایی $\mathcal{D} = \{00, 1, 01\}$ (که همان کدکلمه‌ها را با ترتیب دیگری به‌کار می‌برد) لحظه‌ای بوده و

$$L(\mathcal{C}) = \frac{1}{4}.2 + \frac{1}{2}.1 + \frac{1}{4}.2 = 1.5.$$

پس $L(\mathcal{D}) < L(\mathcal{C})$، و ملاحظه این‌که $\mathcal{D}$ یک کد بهینه دوتایی است کار دشواری نیست، یعنی این‌که برای هر کد دوتایی لحظه‌ای $\mathcal{C}$ برای $\mathcal{S}$ رابطه $L(\mathcal{D}) \leq L(\mathcal{C})$ برقرار است.

---

[1] compact

این مثال یک قاعده کلی را توضیح می‌دهد که براساس آن میانگین طول کد با نسبت دادن کدکلمه‌های کوچک به سمبل‌های با احتمال وقوع زیادتر کاهش می‌یابد (به‌کار بردن $w_۲ = ۱$ در $\mathcal{D}$ به عوض $w_۲ = ۰۱$ در $\mathcal{C}$)؛ کد مورس[۱] از یک چنین ایده‌ای استفاده می‌کند.

تمرین ۱.۲ نشان دهید که در یک کد بهینه اگر $p_i > p_j$ آنگاه $l_i \leq l_j$.

از این اصل در بخش‌های بعد به شکل قانونمندی برای ساخت کدهای بهینه استفاده می‌کنیم. ابتدا نشان می‌دهیم که میانگین طول کد با درنظر گرفتن کدهای یکتا دکدپذیر قابل کاهش بیشتر، در مقایسه با تنها لحاظ نمودن کدهای لحظه‌ای، نیست و از این‌رو اگر تنها کدهای لحظه‌ای را درنظر بگیریم هیچ محدودیتی ایجاد نمی‌شود. لم زیر فوراً از نتیجه ۲۲.۱ به‌دست می‌آید.

لم ۲.۲ منبع S و عدد صحیح r را درنظر بگیرید. مجموعه تمامی میانگین طول کدهای $L(\mathcal{C})$ که یک کد $\mathcal{C}$ یک کد r–تایی یکتا دکدپذیر برای S است برابر است با مجموعه تمامی میانگین طول کدهای $L(\mathcal{C})$ که $\mathcal{C}$ یک کد r–تایی لحظه‌ای برای S است.

این مجموعه از میانگین طول کدها آشکارا از پایین به‌وسیله ۰ محدود می‌شود، از این‌رو فرض کنید $L_{\min}(S)$ معرف بزرگ‌ترین کران پایین برای این مجموعه باشد (مقدار r معلوم فرض می‌شود). یک کد لحظه‌ای $\mathcal{C}$ بهینه تعریف می‌شود اگر $L(\mathcal{C}) = L_{\min}(S)$. وجود چنین کدهایی کاملاً واضح نیست؛ منطقی به‌نظر می‌رسد که تصور شود که کدهای r–تایی لحظه‌ای میانگین طول کدهایی داشته باشند که به این بزرگ‌ترین کران پایین میل کرده ولی مساوی آن نباشند (همچون دنباله $\frac{۱}{n}$، $n \in \mathbf{N}$، با بزرگ‌ترین کران پایین ۰). بنابراین وجود کدهای بهینه نیاز به اثبات دارد (نکته‌ای که معمولاً نادیده گرفته می‌شود).

قضیه ۳.۲ هرمنبع S برای هر $r \geq ۲$ دارای یک کد بهینه r–تایی است.

اثبات با درنظر گرفتن این‌که در صورت نیاز می‌توان سمبل‌های $s_۱, \ldots, s_q$ را شماره‌گذاری مجدد نمود، فرض می‌کنیم عددی چون k موجود است به‌قسمی‌که $p_i > ۰$ اگر $i \leq k$ و $p_i = ۰$ برای $i > k$. فرض کنید $p = \min(p_۱, \ldots, p_k)$، پس $p > ۰$.

قطعاً یک کد r–تایی لحظه‌ای $\mathcal{C}$ برای S وجود دارد؛ به‌عنوان نمونه، می‌توانیم برای عددی چون l با خاصیت $r^l \geq q$ قرار دهیم $l_۱ = \cdots = l_q = l$ و قضیه ۲۰.۱ را به‌کار ببریم. برای اثبات قضیه کافی است نشان داده شود که در میان تمامی کدهای r–تایی لحظه‌ای $\mathcal{D}$ برای S تنها برای تعداد متناهی از آن‌ها خاصیت $L(\mathcal{D}) \leq L(\mathcal{C})$ برقرار است؛ کم‌ترین مقدار از این تعداد متناهی عدد توسط کدی چون $\mathcal{D}$ به‌دست می‌آید که باید در واقع بهینه باشد.

برای نشان دادن این خاصیت، فرض کنید $\mathcal{D}$ یک کد r–تایی لحظه‌ای با شرط $L(\mathcal{D}) \leq L(\mathcal{C})$ باشد. در این‌صورت طول کدهای $l_۱, \ldots, l_q$ از $\mathcal{D}$ باید در شرط زیر صدق

[۱] Morse code

کنند

$$l_i \le \frac{L(\mathcal{C})}{p}, \quad i = 1, \dots, k,$$

زیرا در غیر این‌صورت خواهیم داشت

$$L(\mathcal{D}) = p_1 l_1 + \cdots + p_q l_q \ge p_i l_i > p\frac{L(\mathcal{C})}{p} = L(\mathcal{C}).$$

تنها تعداد متناهی کلمه $w \in T^+$ با خاصیت $|w| \le \frac{L(\mathcal{C})}{p}$ وجود دارد، و از این‌رو برای کدکلمه‌های $w_1, \dots, w_k$ از $\mathcal{D}$ تنها تعداد متناهی انتخاب وجود دارد. برای هر w_i، $k < i$، تعداد نامتناهی انتخاب وجود دارد، ولی آن‌ها تأثیری روی $L(\mathcal{D})$ ندارند زیرا برای این i رابطه $p_i = 0$ برقرار است. در نتیجه تنها تعداد متناهی مقدار $L(\mathcal{D})$ در $L(\mathcal{D}) \le L(\mathcal{C})$ صدق می‌کنند. ■

تمرین ۲.۲ یک تعبیر هندسی از قضیه ۳.۲ ارائه دهید؛ مسئله را می‌توان می‌نیمم کردن ضرب اسکالر $L = \sum p_i l_i = \mathbf{p}.\mathbf{l}$ درنظر گرفت که در آن بردارهای $\mathbf{p} = (p_1, \dots, p_q)$ و $\mathbf{l} = (l_1, \dots, l_q)$ در $\mathbf{R}^q$ بوده و دارای محدودیت‌های خاصی می‌باشند. (راهنمایی: ابتدا حالت $q = 2$ را درنظر بگیرید.)

۲.۲ کدهای دوتایی هافمن

در ۱۹۵۲ هافمن [Hu 52] الگوریتمی برای ساخت کدهای بهینه معرفی کرد. برای سادگی ما روی حالت دوتایی متمرکز خواهیم شد، پس فرض کنید $T = \mathbf{Z}_2 = \{0, 1\}$. برای یک منبع مفروض $\mathcal{S}$، سمبل‌های $s_1, \dots, s_q$ را در صورت لزوم شماره‌گذاری مجدد کرده تا خاصیت $p_1 \ge p_2 \ge \cdots \ge p_q$ برقرار باشد. یک منبع کاهش‌یافته $\mathcal{S}'$ از روی $\mathcal{S}$ با ادغام دو سمبل با کمترین احتمال وقوع، یعنی s_{q-1} و s_q، و معرفی سمبل $s' = s_{q-1} \vee s_q$ (به معنی s_{q-1} یا s_q) تشکیل می‌دهیم که در آن احتمال وقوع s' برابر $p' = p_{q-1} + p_q$ می‌باشد (اگر زوج با کمترین احتمال وقوع منحصربه‌فرد نباشد یک انتخاب دلخواه از دو سمبل با کمترین احتمال را درنظر می‌گیریم). پس $\mathcal{S}'$ منبع با سمبل‌های $s_1, \dots, s_{q-2}, s'$ و توزیع احتمال $p_1, \dots, p_{q-2}, p'$ می‌باشد.

اگر $\mathcal{C}'$ یک کد دوتایی برای $\mathcal{S}'$ باشد می‌توان از روی آن یک کد دوتایی $\mathcal{C}$ برای $\mathcal{S}$ ساخت: اگر $\mathcal{C}'$ سمبل s_i را به w_i، $1 \le i \le q-2$، کد کند آنگاه $\mathcal{C}$ نیز چنین خواهد کرد، و اگر $\mathcal{C}'$ سمبل s' را به w' کد کند آنگاه $\mathcal{C}$ سمبل‌های s_{q-1} و s_q را به‌ترتیب به $w'0$ و $w'1$ کد می‌کند. این عمل به شکل زیر توضیح داده می‌شود:

$$\mathcal{S} \ : \ s_1, \dots, s_{q-2}, \underbrace{s_{q-1}, s_q}$$

$$\begin{aligned}
\mathcal{S}' &: s_1,\ldots,s_{q-2},\quad s'\\
&\quad p_1,\ldots,p_{q-2},\underbrace{p_{q-1},p_q}\\
&\quad p_1,\ldots,p_{q-2},\quad p'\\
\mathcal{C} &: w_1,\ldots,w_{q-2},\underbrace{w'0,w'1}\\
\mathcal{C}' &: w_1,\ldots,w_{q-2},\quad w'
\end{aligned}$$

لم ۴.۲ اگر کد $\mathcal{C}'$ لحظه‌ای باشد آنگاه $\mathcal{C}$ نیز چنین است.

اثبات به‌سادگی می‌توان نشان داد که اگر $\mathcal{C}'$ یک کد پیشوند باشد آنگاه $\mathcal{C}$ نیز چنین خواهد بود؛ اکنون قضیه ۱۷.۱ اثبات را کامل می‌کند. ■

بر این اساس از روی یک کد لحظه‌ای $\mathcal{C}'$ برای $\mathcal{S}'$ (که دارای $q-1$ سمبل است) یک کد لحظه‌ای $\mathcal{C}$ برای $\mathcal{S}$ (با q سمبل) ساخته می‌شود. مشابهاً یک کد لحظه‌ای دوتایی $\mathcal{C}'$ برای $\mathcal{S}'$ از روی یک کد لحظه‌ای دوتایی $\mathcal{C}''$ برای $\mathcal{S}''$ (که دارای $q-2$ سمبل است) نتیجه می‌شود، که $\mathcal{S}''$ از روی $\mathcal{S}'$ با ادغام نمودن دو سمبل با کمترین احتمال به‌دست آمده است. اگر کاهش دادن منابع را به‌همین روش ادامه دهیم دنباله $\mathcal{S}, \mathcal{S}', \ldots, \mathcal{S}^{(q-2)}, \mathcal{S}^{(q-1)}$ از منابع را با تعداد سمبل‌های به‌ترتیب $q, q-1, \ldots, 2, 1$ به‌دست می‌آوریم:

$$\mathcal{S} \longrightarrow \mathcal{S}' \longrightarrow \cdots \longrightarrow \mathcal{S}^{(q-2)} \longrightarrow \mathcal{S}^{(q-1)}.$$

اکنون منبع $\mathcal{S}^{(q-1)}$ دارای تنها سمبل $s_1 \vee \cdots \vee s_q$ با احتمال وقوع ۱ بوده و ما آن‌را با کلمه خالی ε کد کرده و کد $\mathcal{C}^{(q-1)} = \{\varepsilon\}$ را برای $\mathcal{S}^{(q-1)}$ درنظر می‌گیریم.[1] روش اضافه کردن ۰ و ۱ به یک کدکلمه w' منجر به ایجاد یک کد لحظه‌ای دوتایی $\mathcal{C}^{(q-2)} = \{\varepsilon 0 = 0, \varepsilon 1 = 1\}$ برای $\mathcal{S}^{(q-2)}$ شده، و با ادامه این روش، به تعداد $q-1$ دفعه، یک دنباله از کدهای دوتایی $\mathcal{C}^{(q-1)}, \mathcal{C}^{(q-2)}, \ldots, \mathcal{C}', \mathcal{C}$ برای منابع $\mathcal{S}^{(q-1)}, \mathcal{S}^{(q-2)}, \ldots, \mathcal{S}', \mathcal{S}$ ساخته می‌شود:

$$\begin{aligned}
&\mathcal{S} \longrightarrow \mathcal{S}' \longrightarrow \cdots \longrightarrow \mathcal{S}^{(q-2)} \longrightarrow \mathcal{S}^{(q-1)}\\
&\mathcal{C} \longleftarrow \mathcal{C}' \longleftarrow \cdots \longleftarrow \mathcal{C}^{(q-2)} \longleftarrow \mathcal{C}^{(q-1)}.
\end{aligned}$$

آخرین کد، یعنی $\mathcal{C}$، یک کد هافمن برای $\mathcal{S}$ نامیده می‌شود. با کاربرد مکرر لم ۴.۲ ملاحظه می‌شود که $\mathcal{C}$ لحظه‌ای است. بهینه بودن $\mathcal{C}$ را در بخش ۴.۲ نشان خواهیم داد. (توجه کنید که هر $\mathcal{C}^{(i)}$ یک کد هافمن برای $\mathcal{S}^{(i)}$ است، زیرا می‌توان برای تمامی $i < j$ از $\mathcal{S}^{(j)}$ و $\mathcal{C}^{(j)}$ صرف‌نظر کرد.)

[1] با بیانی دقیق، $\mathcal{C}^{(q-1)}$ یک کد نیست زیرا تنها حاوی کلمه خالی است؛ ولی نکته این است که امکان ساخت کدهای $\mathcal{C}^{(q-2)}, \ldots, \mathcal{C}$ را فراهم می‌کند.

مثال ۵.۲ فرض کنید $\mathcal{S}$ پنج سمبل s_1، ...، s_5 با احتمال‌های ۰.۳، ۰.۲، ۰.۲، ۰.۲، ۰.۱ داشته باشد. ابتدا عمل کاهش منبع را اجرا می‌کنیم؛ توزیع‌های احتمال متناظر با منابع کاهش‌یافته به شرح زیر است:

$$\begin{array}{lccccccc}
\mathcal{S}: & & & ۰.۳ & & ۰.۲ & ۰.۲ & ۰.۲ & ۰.۱ \\
\mathcal{S}': & & & ۰.۳ & \overline{۰.۳} & ۰.۲ & ۰.۲ & & \\
\mathcal{S}: & & \overline{۰.۴} & ۰.۳ & ۰.۳ & & & & \\
\mathcal{S}''': & & \overline{۰.۶} & ۰.۴ & & & & & \\
\mathcal{S}^{(۴)}: & \overline{۱} & & & & & & &
\end{array}$$

با درنظر گرفتن یک سطر، سطر بعدی به این‌صورت تشکیل می‌شود که دو احتمال با کمترین مقدار با مجموع آن‌ها جایگزین شده و در موقعیتی قرار داده می‌شود که توزیع احتمال جدید از راست به چپ صعودی باشد. اکنون این روند را معکوس کرده و برای این منابع کد می‌سازیم؛ از پایین با ε شروع کرده و با ادامه کار به سمت بالا کدها ساخته می‌شوند:

$$\begin{array}{lccccccc}
\mathcal{C}: & & & ۰۰ & & ۱۰ & ۱۱ & ۰۱۰ & ۰۱۱ \\
\mathcal{C}': & & & ۰۰ & \overline{۰۱} & ۱۰ & ۱۱ & & \\
\mathcal{C}': & & \overline{۱} & ۰۰ & ۰۱ & & & & \\
\mathcal{C}''': & \overline{۰} & ۱ & & & & & & \\
\mathcal{C}^{(۴)}: & \overline{\varepsilon} & & & & & & &
\end{array}$$

هر سطر از سطر زیردست خودش با انتخاب کدکلمه w'، متناظر با سمبل جدید s' در سطر زیری، و اضافه کردن ۰ و ۱ به انتهای آن و نسبت دادن کلمات w'۰ و w'۱ به دو سمبل تشکیل دهنده s' به‌دست می‌آید؛ بقیه کدکلمات سطر زیری بدون تغییر به سطر بالا منتقل می‌شوند. نهایتاً کد دوتایی $\mathcal{C} = \{۰۰, ۱۰, ۱۱, ۰۱۰, ۰۱۱\}$ برای $\mathcal{S}$ حاصل می‌شود. به وضوح این یک کد پیشوند و بنابراین لحظه‌ای است. طول کدهای آن برابر $l_i = ۲, ۲, ۲, ۳, ۳$ بوده و بنابراین

$$L(C) = \sum p_i l_i = ۰.۶ + ۰.۴ + ۰.۴ + ۰.۶ + ۰.۳ = ۲.۳.$$

در اکثر حالات، پروسه کاهش منحصربه‌فرد بوده، و بنابراین کد هافمن یکتا است، به‌جز این‌که در هر مرحله نسبت دادن کلمات w'۰ و w'۱ یا w'۱ و w'۰ به دو سمبل با کمترین احتمال مربوطه اختیاری است. بنابه قرارداد معمولاً حالت اول درنظر گرفته می‌شود اگرچه این انتخاب تأثیری روی طول کدکلمه‌ها ندارد. در بعضی حالات ممکن است در بعضی از مراحل بیش از دو سمبل با کمترین احتمال موجود باشد و از این‌رو ممکن است پروسه کاهش یکتا نباشد، که این باعث وجود انتخاب‌های بیشتری برای کد هافمن می‌شود. این وضع در مثال فوق در مرحله یک اتفاق افتاده، و تمرین ۳.۲ نیز این رفتار را توضیح می‌دهد.

با این همه، بهینگی کد هافمن (که بعداً ثابت می‌شود) ایجاب می‌کند که در یک چنین حالاتی $L(\mathcal{C})$ در هر حال منحصربه‌فرد است.

تمرین ۳.۲ یک کد دوتایی هافمن برای منبع با احتمال‌های

$$p_i = ۰.۴, ۰.۳, ۰.۱, ۰.۱, ۰.۰۶, ۰.۰۴,$$

ساخته و متوسط طول کد را تعیین کنید. تا چه حدی کد، طول کدکلمه‌ها، و میانگین طول کد یکتا هستند؟

براساس ساخت کد هافمن هر طول کد l_i برابر تعداد دفعاتی است که سمبل متناظر آن s_i در S تحت روند کاهش ادغام شده است. این بدین جهت است که ساخت w_i از ε، به‌طول ۰، شروع شده و یک سمبل ۰ یا ۱ اضافه می‌شود دقیقاً وقتی‌که s_i ادغام می‌گردد. پس سمبل‌های با احتمال بالا به دلیل این‌که کمتر ادغام می‌شوند متناظر با کدکلمه‌های کوتاه می‌باشند، که این نکته خود کمک به بهینه بودن کد هافمن می‌کند (اگرچه آن را اثبات نمی‌کند).

برای این‌که تاثیر توزیع احتمال را در ساخت کد هافمن ببینیم به مثال دیگری می‌پردازیم.

مثال ۶.۲ فرض کنید S دارای $q = ۵$ سمبل بوده و احتمال‌ها برابر باشند، یعنی $p_۱ = \cdots = p_۵ = ۰.۲$. روند کاهش و کدگذاری به فرم زیر است:

$\mathcal{S}$:				۰.۲	۰.۲	۰.۲	۰.۲	۰.۲
$\mathcal{S}'$:			$\overline{۰.۴}$	۰.۲	۰.۲	۰.۲		
$\mathcal{S}$:		$\overline{۰.۴}$	۰.۴	۰.۲				
$\mathcal{S}'''$:	$\overline{۰.۶}$	۰.۴						
$\mathcal{S}^{(۴)}$:	$\overline{۱}$							

$\mathcal{C}$:				۰۱	۱۰	۱۱	۰۰۰	۰۰۱
$\mathcal{C}'$:			$\overline{۰۰}$	۰۱	۱۰	۱۱		
$\mathcal{C}'$:		$\overline{۱}$	۰۰	۰۱				
$\mathcal{C}'''$:	$\overline{۰}$	۱						
$\mathcal{C}^{(۴)}$:	$\overline{\varepsilon}$							

از این‌جا یک کد هافمن $\mathcal{C} = \{۰۱, ۱۰, ۱۱, ۰۰۰, ۰۰۱\}$ برای S حاصل می‌شود؛ این کد پیشوند و بنابراین لحظه‌ای است. طول کدهای آن $l_i = ۲, ۲, ۲, ۳, ۳$ بوده و از این‌رو متوسط طول کد برابر ۲.۴ است:

$$L(\mathcal{C}) = \frac{۱}{۵}(۲ + ۲ + ۲ + ۳ + ۳) = ۲.۴.$$

این کمی بزرگ‌تر از مقدار ۲.۳ است که در مثال ۵.۲ به‌دست آمد، که در آن‌جا سمبل‌ها با احتمال یکسان نبودند.

در حالت کلی، هر چه تغییرات بین احتمال‌های p_i بیشتر باشد متوسط طول کد یک کد بهینه پایین‌تر خواهد بود، زیرا دامنه وسیعتری برای نسبت دادن کدکلمه‌های کوتاه به سمبل‌های با احتمال بیشتر وجود دارد. این پدیده را ما در فصل ۳ با استفاده از مفهوم آنتروپی، که میزان تغییرات در توزیع احتمال را اندازه‌گیری می‌کند، مطالعه خواهیم کرد.

تمرین ۴.۲ یک منبع سه سمبل با احتمال‌های $p_1 \geq p_2 \geq p_3$ دارد. نشان دهید یک کد دوتایی هافمن برای این منبع دارای متوسط طول $p_1 - 2$ است. نتیجه مشابهی برای یک منبع با احتمال‌های $p_1 \geq p_2 \geq p_3 \geq p_4$ پیدا کنید.

۳.۲ متوسط طول کلمه در کدهای هافمن

اجازه دهید به وضعیت کلی بخش ۲.۲ بازگشته و متوسط طول کدکلمه در کدهای C و C' را مقایسه کنیم. در C' سمبل $s' = s_{q-1} \vee s_q$ دارای احتمال $p' = p_{q-1} + p_q$ بوده و متناظر با کدکلمه w' است؛ فرض کنید $l = |w'|$. در C سمبل s' با دو سمبل s_{q-1} و s_q، با احتمال‌های p_{q-1} و p_q، جایگزین شده و این دو سمبل با دو کلمه $w'0$ و $w'1$، به‌طول $l+1$، معرفی می‌شوند. هر یک از سمبل‌های $s_1, \ldots, s_{q-2}$ در C و C' کدکلمه‌های مشابه‌ای دارند و از این‌رو

$$\begin{aligned} L(C) - L(C') &= p_{q-1}(l+1) + p_q(l+1) - (p_{q-1} + p_q)l \\ &= p_{q-1} + p_q \\ &= p', \end{aligned} \tag{۳.۲}$$

که احتمال جدید منتج از کاهش S به S' می‌باشد. با تکرار رابطه و استفاده از $L(C^{(q-1)}) = |\varepsilon| = 0$ ملاحظه می‌کنیم که

$$\begin{aligned} L(C) &= (L(C) - L(C')) + (L(C') - L(C'')) + \cdots \\ &\quad \cdots + (L(C^{(q-2)}) - L(C^{(q-1)})) + L(C^{(q-1)}) \\ &= (L(C) - L(C')) + (L(C') - L(C'')) + \cdots + (L(C^{(q-2)}) - L(C^{(q-1)})) \\ &= p' + p'' + \cdots + p^{(q-1)}. \end{aligned} \tag{۴.۲}$$

این عدد مجموع تمامی احتمال‌های جدید p'، p''، ...، $p^{(q-1)}$ است که از کاهش S به $S^{(q-1)}$ نتیجه می‌شوند. به‌عنوان نمونه در مثال ۵.۲ (بخش ۲.۲) احتمال‌های باردار شده را

جمع نموده ونتیجه می‌گیریم

$$L(\mathcal{C}) = ۰.۳ + ۰.۴ + ۰.۶ + ۱ = ۲.۳,$$

در حالی‌که در مثال ۶.۲ داریم

$$L(\mathcal{C}) = ۰.۴ + ۰.۴ + ۰.۶ + ۱ = ۲.۴.$$

این روش خوبی برای کاهش عملیات لازم می‌باشد، زیرا بدون ساخت $\mathcal{C}$ اجازه محاسبه $L(\mathcal{C})$ را می‌دهد: به‌عنوان نمونه در مثال ۵.۲ از توزیع احتمال $p_i = ۰.۳, ۰.۲, ۰.۲, ۰.۲, ۰.۱$ واضح است که با ادغام متوالی کوچک‌ترین زوج‌ها باید داشته باشیم

$$\begin{aligned} p' &= ۰.۲ + ۰.۱ = ۰.۳, \\ p'' &= ۰.۲ + ۰.۲ = ۰.۴, \\ p''' &= ۰.۳ + ۰.۳ = ۰.۶, \\ p'''' &= ۰.۴ + ۰.۶ = ۱, \end{aligned}$$

و بنابراین $L(\mathcal{C}) = ۰.۳ + ۰.۴ + ۰.۶ + ۱ = ۲.۳$.

تمرین ۵.۲ با به‌کار بردن این روش درستی متوسط طول کدهای تمرینات ۳.۲ و ۴.۲ را تصدیق کنید.

۴.۲ بهینگی کدهای هافمن دوتایی

در این بخش نشان می‌دهیم که کدهای هافمن دوتایی بهینه هستند. ابتدا به یک تعریف و لم نیاز داریم. دو کلمه دوتایی $w_۱$ و $w_۲$ را هم‌زاد[۱] نامیم هر گاه برای عضوی چون $x \in T^*$ این دو به شکل $x۰$ و $x۱$ (یا به‌عکس) باشند.

لم ۷.۲ هر منبع $\mathcal{S}$ دارای یک کد دوتایی بهینه $\mathcal{D}$ است که در آن دو کدکلمه با طول ماکزیمم هم‌زاد وجود دارند.

اثبات بنابه قضیه ۳.۲ یک کد بهینه دوتایی برای $\mathcal{S}$ وجود دارد؛ در میان تمامی چنین کدهایی فرض کنید $\mathcal{D}$ کدی باشد که $\sigma(\mathcal{D}) = \sum_i l_i$ را می‌نیمم می‌کند (به‌جهت این‌که طول کلمات اعداد صحیح نامنفی هستند چنین کدی موجود است). نشان می‌دهیم که $\mathcal{D}$ خاصیت لازم را دارد.

یک کلمه با طول ماکزیمم در $\mathcal{D}$ انتخاب کنید؛ این باید به‌فرم xt باشد که $x \in T^*$ و $t \in T = \mathbf{Z}_۲$. فرض کنید $\bar{t}$ معرف $t - ۱$ باشد، پس $\bar{t} = ۰$ یا $\bar{t} = ۱$. اگر $x\bar{t}$ در $\mathcal{D}$ باشد

[۱] siblings

آنگاه xt و $x\bar{t}$ هم‌زاد مطلوب هستند؛ پس فرض کنید که $x\bar{t} \notin \mathcal{D}$. به‌جهت لحظه‌ای بودن، $\mathcal{D}$ یک کد پیشوند است. اکنون تنها کدکلمه با پیشوند x کلمه xt است (زیرا $|xt|$ ماکزیمم بوده و $x\bar{t} \notin \mathcal{D}$)، پس اگر کلمهٔ xt را با x تعویض کنیم کد جدید $\mathcal{D}'$ را برای S به‌دست می‌آوریم که پیشوند نیز می‌باشد. پس $\mathcal{D}'$ یک کد لحظه‌ای برای S با خاصیت $L(\mathcal{D}') \leq L(\mathcal{D})$ بوده و $\sigma(\mathcal{D}') = \sigma(\mathcal{D}) - ۱ < \sigma(\mathcal{D})$، که در تناقض با انتخاب $\mathcal{D}$ است. پس $x\bar{t} \in \mathcal{D}$. ■

قضیه ۸.۲ اگر $\mathcal{C}$ یک کد دوتایی هافمن برای منبع S باشد، آنگاه $\mathcal{C}$ یک کد بهینه برای S است.

اثبات بنابر لم ۴.۲ کد $\mathcal{C}$ لحظه‌ای است، پس کافی است نشان داده شود که $L(\mathcal{C})$ می‌نیمال است (در میان متوسط طول کد تمامی کدهای دوتایی لحظه‌ای برای S). از استقرا روی تعداد سمبل‌های منبع، q، استفاده می‌کنیم. اگر $q = ۱$ آنگاه $\mathcal{C} = \{\varepsilon\}$ با $L(\mathcal{C}) = ۰$، و از این‌رو در این حالت حکم بدیهی است. پس فرض کنید $q > ۱$ و فرض کنید حکم برای تمامی منابع با $q - ۱$ سمبل برقرار باشد.

فرض کنید S' منبع به‌دست آمده از کاهش S، مطابق بخش ۲.۲، باشد، پس S' دارای $q - ۱$ سمبل $s_۱, s_۲, \ldots, s_{q-۲}, s_{q-۱} \vee s_q$ است. بنابر (۳.۲) داریم $L(\mathcal{C}) - L(\mathcal{C}') = p_{q-۱} + p_q = p'$، که احتمال s' است.

حال فرض کنید $\mathcal{D}$ یک کد دوتایی بهینه برای S باشد که مطابق با لم ۷.۲ به‌دست آمده است و در آن s_i متناظر با کلمه x_i است. پس $\mathcal{D}$ حاوی یک زوج هم‌زاد از بزرگ‌ترین کلمات $x_u = x۰$، $x_v = x۱$ متناظر با سمبل‌های s_u و s_v می‌باشد. نشان می‌دهیم که می‌توان فرض کرد $u = q - ۱$ و $v = q$.

اگر $v \neq q$ آنگاه می‌توان کلمات x_v و x_q را جابه‌جا نموده و یک کد لحظه‌ای $\mathcal{D}'$ برای S به‌دست آورد؛ اگر m_i معرف $|x_i|$ باشد آنگاه این جابه‌جایی عبارت $p_v m_v + p_q m_q$ در $L(\mathcal{D})$ را با $p_v m_q + p_q m_v$ در $L(\mathcal{D}^*)$ جایگزین می‌کند. حال از $p_v \geq p_q$ و $m_v \geq m_q$ داریم:

$$(p_v m_v + p_q m_q) - (p_v m_q + p_q m_v) = (p_v - p_q)(m_v - m_q) \geq ۰,$$

پس $L(\mathcal{D}) \geq L(\mathcal{D}^*)$. از بهینه بودن $\mathcal{D}$ نتیجه می‌شود $L(\mathcal{D}) = L(\mathcal{D}^*)$ و بنابراین $\mathcal{D}^*$ بهینه است. چون در صورت لزوم می‌توان $\mathcal{D}$ را با $\mathcal{D}^*$ جایگزین کرد، می‌توان فرض کرد که $v = q$. بحث مشابهی نشان می‌دهد که می‌توان فرض کرد $u = q - ۱$، و از این‌رو هم‌زادهای $x۰$ و $x۱$ در $\mathcal{D}$ کدکلمات متناظر با $s_{q-۱}$ و s_q هستند.

حال یک کد $\mathcal{D}'$ برای S' تشکیل می‌دهیم که با $s_i \longmapsto x_i$، $۱ \leq i \leq q - ۲$ و $s' \longmapsto x$ معرفی می‌شود. پس رابطه $\mathcal{D}$ و $\mathcal{D}'$ مانند رابطه $\mathcal{C}$ و $\mathcal{C}'$ است. بالاخص، با به‌کار بردن بحث

بخش ۳.۲ روی $\mathcal{D}$ و $\mathcal{D}'$ ملاحظه می‌شود که

$$L(\mathcal{D}) - L(\mathcal{D}') = p_{q-1} + p_q = L(\mathcal{C}) - L(\mathcal{C}'),$$

و از این‌رو $L(\mathcal{D}') - L(\mathcal{C}') = L(\mathcal{D}) - L(\mathcal{C})$.

حال $\mathcal{C}'$ یک کد هافمن برای منبع $\mathcal{S}'$، با $q-1$ سمبل، بوده و بنابه فرض استقرا $\mathcal{C}'$ بهینه است. پس $L(\mathcal{D}') \leq L(\mathcal{C}')$ و از این‌رو $L(\mathcal{D}) \leq L(\mathcal{C})$. چون $\mathcal{D}$ بهینه است $\mathcal{C}$ نیز چنین است و $L(\mathcal{C}) = L(\mathcal{D})$. ■

تمرین ۶.۲ روی این استدلال نظر بدهید: هر منبع یک کد هافمن دارد؛ تمامی کدهای هافمن بهینه هستند؛ پس هر منبع یک کد بهینه دارد.

۵.۲ کدهای هافمن r–تایی

اگر از یک الفبای T با $|T| = r > 2$ استفاده کنیم آنگاه ساخت کدهای هافمن r–تایی مشابه حالت دوتایی خواهد بود. اگر $\mathcal{S}$ یک منبع باشد، یک دنباله از منابع کاهش یافته $\mathcal{S}'$، $\mathcal{S}''$، ...، تشکیل می‌دهیم که در آن هر بار r سمبل با کمترین احتمال ادغام شده با سمبل s' معرفی گردیده، و احتمال s' برابر مجموع احتمال‌های r سمبل ادغام شده تعریف می‌شود.

می‌خواهیم نهایتاً $\mathcal{S}$ را به یک منبع یک عضوی با احتمال ۱ کاهش دهیم، که با کدکلمه ε معرفی می‌شود. چون در هر مرحله تعداد سمبل‌ها به اندازه $r-1$ کاهش می‌یابد این کار ممکن است اگر و فقط اگر $q \equiv 1 \bmod (r-1)$. این شرط برای حالت $r = 2$ همیشه برقرار است، ولی لزوماً برای $r > 2$ برقرار نیست. اگر $q \neq 1 \bmod (r-1)$، به تعداد کافی سمبل‌های اضافی s_i با احتمال $p_i = 0$ به $\mathcal{S}$ ضمیمه می‌کنیم تا این هم‌نهشتی برقرار شده و سپس روند کاهش را اجرا می‌کنیم.

مثال ۹.۲ فرض کنید $q = 6$ و $r = 3$. چون $r - 1 = 2$ نیاز به $q = 1 \bmod 2$ داریم و بنابراین یک سمبل اضافی s_7 با احتمال $p_7 = 0$ به $\mathcal{S}$ ضمیمه می‌کنیم. حال روند کاهش منابع $\mathcal{S}'$، $\mathcal{S}''$ و $\mathcal{S}'''$ را به‌ترتیب با ۵، ۳ و ۱ سمبل ایجاد می‌کند.

ساخت کد $\mathcal{C}$ شبیه حالت دوتایی است. اگر $\mathcal{C}^{(i)}$ یک کد برای $\mathcal{S}^{(i)}$ باشد، یک کد $\mathcal{C}^{(i-1)}$ را به روش زیر برای $\mathcal{S}^{(i-1)}$ تشکیل می‌دهیم: این عمل با حذف نمودن کدکلمه w'، متناظر با سمبل جدید $s' \in \mathcal{S}^{(i)}$، و جایگزین کردن آن با r کدکلمه $(t \in T)w't$ برای r سمبل از منبع $\mathcal{S}^{(i-1)}$ که به s' ادغام شده‌اند صورت می‌گیرد. با تکرار این عمل، نهایتاً یک کد هافمن r–تایی $\mathcal{C}$ برای $\mathcal{S}$ حاصل می‌شود که البته با حذف کدکلمات متناظر با سمبل‌های اضافی s_i که در ابتدا به منبع اضافه شده بودند همراه می‌باشد.

مثال ۱۰.۲ همچون مثال ۹.۲ فرض کنید $q = ٦$ و $r = ٣$، و فرض کنید که سمبل‌های $s_١, \ldots, s_٦$ از $\mathcal{S}$ توزیع احتمال $p_i = (٠.٣, ٠.٢, ٠.٢, ٠.١, ٠.١, ٠.١)$ داشته باشند. بعد از ضمیمه نمودن $s_٧$ با $p_٧ = ٠$ روند کاهش را به‌شکل زیر خواهیم داشت:

$$
\begin{array}{rcccccccc}
\mathcal{S}: & & ٠.٣ & ٠.٢ & ٠.٢ & \_\_ & ٠.١ & ٠.١ & ٠.١ & ٠ \\
\mathcal{S}': & & ٠.٣ & ٠.٢ & ٠.٢ & \overline{٠.٢} & ٠.١ & & & \\
\mathcal{S}'': & \overline{٠.٥} & ٠.٣ & ٠.٢ & & & & & & \\
\mathcal{S}''': & \overline{١} & & & & & & & &
\end{array}
$$

اگر قرار دهیم $T = \mathbf{Z}_٣ = \{٠, ١, ٢\}$، آنگاه یک امکان کدگذاری به‌فرم زیر است:

$$
\begin{array}{rcccccccc}
\mathcal{C}: & & ١ & ٢ & ٠٠ & \_\_ & ٠٢ & ٠١٠ & ٠١١ & ٠١٢ \\
\mathcal{C}': & & ١ & ٢ & ٠٠ & \overline{٠١} & ٠٢ & & & \\
\mathcal{C}'': & \overline{٠} & ١ & ٢ & & & & & & \\
\mathcal{C}''': & \overline{\varepsilon} & & & & & & & &
\end{array}
$$

با حذف کلمه ٠١٢، متناظر با سمبل $s_٧$، کد سه‌تایی هافمن $\mathcal{C}$ را، با $L(\mathcal{C}) = ١.٧$، برای $\mathcal{S}$ به‌دست می‌آوریم که $\mathcal{C} = \{١, ٢, ٠٠, ٠٢, ٠١٠, ٠١١\}$.

اثبات این‌که کدهای هافمن r–تایی لحظه‌ای هستند شبیه به حالت $r = ٢$ است؛ با این حال اثبات بهینگی کمی دشوارتر از حالت دوتایی است زیرا لم ۷.۲ به آسانی قابل تعمیم به‌حالت $r > ٢$ نیست، از این‌رو ما آن را حذف می‌کنیم. اثبات (۴.۲)، یعنی این‌که $L(\mathcal{C})$ برابر مجموع $p' + p'' + \cdots$ از احتمال‌های جدید است برای حالت غیر دوتایی نیز برقرار است؛ به‌عنوان نمونه در مثال ۱۰.۲ تساوی $L(\mathcal{C}) = ٠.٢ + ٠.٥ + ١ = ١.٧$ برقرار است.

تمرین ۷.۲ برای منبع با توزیع احتمال زیر کدهای هافمن دوتایی و سه‌تایی پیدا کنید:

$$p_i = ٠.٣, ٠.٢, ٠.١٥, ٠.١, ٠.١, ٠.٠٨, ٠.٠٥, ٠.٠٢.$$

در هر یک از حالات میانگین طول کد را محاسبه کنید.

تمرین ۸.۲ رابطه (۴.۲) را به حالت r–تایی با $r > ٢$ تعمیم دهید.

٦.٢ بسط منابع

به‌جای کدکردن سمبل‌های s_i، یک سمبل با یک کلمه، کاراتر خواهد بود اگر قالبی از سمبل‌های متوالی با یک کلمه کد شود، به‌عنوان مثال بهتر است کلمات (یا حتی جملات) یک متن به‌عوض حروف کدگذاری شوند. این امر باعث تغییرات بیشتر در احتمال‌ها شده و

از این‌رو امکان داشتن متوسط طول کد کمتری را فراهم می‌کند (همچنان‌که در بخش ۲.۲ اشاره شد).

فرض کنید $\mathcal{S}$ یک منبع با الفبای S شامل q سمبل $s_1, \ldots, s_q$ و توزیع احتمال $p_1, \ldots, p_q$ باشد. بسط n ام $\mathcal{S}$ منبع با الفبای S^n شامل q^n سمبل $s_{i_1} \ldots s_{i_n}$ $(s_{i_j} \in S)$ و توزیع احتمال $p_{i_1} \ldots p_{i_n}$ می‌باشد. یک سمبل $s_{i_1} \ldots s_{i_n} \in \mathcal{S}^n$ را می‌توان به‌عنوان یک قالب از n سمبل متوالی از $\mathcal{S}$، یا معادلاً به‌عنوان یک خروجی از n کپی مستقل از $\mathcal{S}$ (تصور کنید پرتاب چند سکه مشابه و یا چند تاس مشابه را) درنظر گرفت. با بسط سمت چپ معادله

$$(p_1 + \cdots + p_q)^n = 1^n = 1,$$

و ملاحظه این‌که هر $p_{i_1} \ldots p_{i_n}$ یک‌بار ظاهر می‌شود می‌توان دید که احتمال‌های $p_{i_1} \ldots p_{i_n}$ تشکیل یک توزیع احتمال می‌دهند.

مثال ۱۱.۲ فرض کنید $\mathcal{S}$ دارای الفبای منبع $S = \{s_1, s_2\}$ با $p_1 = \frac{2}{3}$ و $p_2 = \frac{1}{3}$ باشد. دراین‌صورت $\mathcal{S}^2$ دارای الفبای $S^2 = \{s_1s_1, s_1s_2, s_2s_1, s_2s_2\}$ با احتمال‌های $\frac{4}{9}$، $\frac{2}{9}$، $\frac{2}{9}$، $\frac{1}{9}$ است.

در حالت کلی فرض کنید p_1 و p_q به‌ترتیب بزرگ‌ترین و کوچک‌ترین احتمال برای $\mathcal{S}$ باشند، بنابراین p_1^n و p_q^n بزرگ‌ترین و کوچک‌ترین احتمال‌ها برای $\mathcal{S}^n$ هستند. با فرض $p_1 > p_q$ (یعنی این‌که تمامی احتمال‌ها برابر $\frac{1}{q}$ نیستند) و $n \longrightarrow \infty$ داریم

$$\frac{p_1^n}{p_q^n} = \left(\frac{p_1}{p_q}\right)^n \longrightarrow \infty.$$

این بدین معنی است که $\mathcal{S}^n$ با افزایش n دارای تغییرات بیشتری در توزیع احتمال خود می‌باشد و از این‌رو می‌توان انتظار کدگذاری کاراتری را داشت.

مثال ۱۲.۲ اگر $\mathcal{S}$ منبع مثال ۱۱.۲ باشد، یک کد دوتایی هافمن C ($s_1 \longmapsto 0$، $s_2 \longmapsto 1$) با میانگین طول کد $L(C) = 1$ وجود دارد. قبول این‌که بتوان این کد را بهبود بخشید دشوار است، با این حال اجازه دهید یک کد هافمن برای $\mathcal{S}^2$ بسازیم. از الگوریتم توصیف شده در بخش ۲.۲ به‌شرح زیر استفاده می‌کنیم (از زحمت دوباره نویسی احتمال‌ها به‌صورت نزولی در هر سطر اجتناب کرده‌ایم):

$\mathcal{S}^2$:	۴/۹	۲/۹	۲/۹	۱/۹	۰	۱۰	۱۱۰	۱۱۱
$(\mathcal{S}^2)'$:	۴/۹	۲/۹	۳/۹		۰	۱۰	۱۱	
$(\mathcal{S}^2)''$:	۴/۹	۵/۹			۰	۱		
$(\mathcal{S}^2)'''$:	۱				ε			

این یک کد هافمن $\mathcal{C}^2$ معرفی می‌کند که در آن $s_1s_1 \longmapsto ۰$، $s_1s_2 \longmapsto ۱۰$، $s_2s_1 \longmapsto ۱۱۰$، $s_2s_2 \longmapsto ۱۱۱$، و میانگین طول کد برابر است با

$$L_2 = L(\mathcal{C}^2) = \frac{۳}{۹} + \frac{۵}{۹} + ۱ = \frac{۱۷}{۹}.$$

هر کدکلمه در $\mathcal{C}^2$ معرف یک قالب دو سمبلی از S است و از این‌رو هر سمبل S به‌طور متوسط به $\frac{۱۷}{۱۸}$ رقم دوتایی نیاز دارد. بنابراین، به‌عنوان یک روش کدگذاری برای S، کد $\mathcal{C}^2$ دارای میانگین طول کلمه $\frac{L_2}{۲}$ است

$$\frac{L_2}{۲} = \frac{۱۷}{۱۸} = ۰.۹۴۴\ldots.$$

این کمتر از متوسط طول کد $L(\mathcal{C}) = ۱$ برای کد هافمن $\mathcal{C}$ بوده و از این‌رو این روش کدگذاری کاراتر است.

به بیان دقیق، آنچه که توصیف کرده‌ایم یک کد برای S نیست زیرا سمبل‌های S کدکلمات خاص خود را ندارند؛ با این وجود، این روش امکان کدگذاری اطلاعاتی را که از منبع S خارج می‌شود فراهم می‌آورد، و بنابراین آن را یک کدگذاری S می‌نامیم. بدین لحاظ این یک کد یکتا دکدپذیر است: به‌عنوان یک کد برای S^2، کد هافمن $\mathcal{C}^2$ لحظه‌ای و بنابراین یکتا دکدپذیر است؛ این بدین معنی است که هر کددنباله t را به‌طور یکتا می‌توان به کدکلمه‌ها تجزیه کرده و سمبل‌های $s_{i_1}s_{i_2}$ از S^2 را به‌دست آورده، سپس سمبل‌های متعلق به S را که در t کد شده‌اند مشخص نمود. ولی این کدگشایی کاملاً لحظه‌ای نیست؛ ما باید سمبل‌های S را در قالب‌های دوتایی کدگشایی کنیم و بنابراین یک تأخیر محدود مادامی‌که منتظر تکمیل یک زوج هستیم ایجاد می‌شود.

با ادامه این روش می‌توان دید که یک کد هافمن $\mathcal{C}^3$ برای S^3 دارای متوسط طول کد $L_3 = L(\mathcal{C}^3) = \frac{۷۶}{۲۷}$ است (تمرین ۹.۲)؛ به‌عنوان یک روش کدگذاری برای S میانگین طول کد

$$\frac{L_3}{۳} = \frac{۷۶}{۸۱} = ۰.۹۳۸\ldots,$$

می‌باشد که حتی بهتر از $\mathcal{C}^2$ است.

این ایده به‌سادگی به S^n تعمیم داده شده و در این رابطه دو سؤال طبیعی ظهور می‌کنند: این‌که وضعیت میانگین طول کد $\frac{L_n}{n}$ وقتی n به بینهایت میل می‌کند چه می‌شود، که در آن $L_n = L(\mathcal{C}^n)$، و این‌که آیا می‌توانیم این روش را برای منابع دیگر به‌کار برده و کدهای کاراتری برای آن‌ها بسازیم. برای جواب به این سؤالات به موضوع اصلی بعدی که آنتروپی است نیاز داریم.

تمرین ۹.۲ فرض کنید S منبع مثال‌های ۱۱.۲ و ۱۲.۲ باشد. توزیع احتمال منبع $S^۳$ را پیدا کنید و نشان دهید که یک کد دوتایی هافمن $C^۳$ برای این منبع دارای متوسط طول کد $L_۳ = L(C^۳) = \frac{۷۶}{۲۷}$ است.

۷.۲ تمرینات تکمیلی

تمرین ۱۰.۲ فرض کنید S یک منبع با احتمال‌های ۰.۳، ۰.۳، ۰.۲، ۰.۲ باشد؛ این منبع چند کد دوتایی بهینه دارد؟ آیا همه آن‌ها کد هافمن هستند؟

تمرین ۱۱.۲ یک منبع S دارای سمبل‌های $s_۱, \ldots, s_q$ با احتمال‌های $p_۱ \geq \cdots \geq p_q$ بوده به‌قسمی‌که برای $i = ۱, \ldots, q-۳$ رابطه $p_i > p_{i+۲} + \cdots + p_q$ برقرار است. ثابت کنید که در هر کد دوتایی هافمن برای S طول کلمات برابر $۱, ۲, \ldots, q-۱, q-۱$ هستند. چند کد دوتایی هافمن متمایز برای S وجود دارد؟ برای هر $q \geq ۱$، مثالی از یک توزیع احتمال (p_i) ارائه دهید که در این نامساوی‌ها صدق کند.

تمرین ۱۲.۲ چگونه می‌توان کدگذاری r–تایی هافمن را پیاده‌سازی کرد به‌قسمی‌که کد هافمن حاصل مجموع طول کد $\sigma(C) = \sum_i l_i$ را نیز می‌نیمم کند؟ (راهنمایی: ابتدا مسئله را برای کد دوتایی هافمن برای توزیع $\frac{۱}{۳}$، $\frac{۱}{۳}$، $\frac{۱}{۹}$، $\frac{۱}{۹}$ حل کنید.)

تمرین ۱۳.۲ یک عضو غیرمعلوم s از مجموعه متناهی $S = \{s_۱, \ldots, s_q\}$ با توزیع احتمال (p_i) انتخاب می‌شود. عضو انتخاب‌شده باید با یک دنباله از سؤالات $Q_۱$، $Q_۲$، ... به‌فرم 'آیا s در T قرار دارد؟' (همچون بازی بیست سؤالی) مشخص شود. روشی برای می‌نیمم کردن تعداد سؤالات لازم ارائه دهید.

تمرین ۱۴.۲ فرض کنید C یک کد دوتایی هافمن برای یک منبع S با توزیع احتمال یکنواخت باشد. آیا امکان این وجود دارد که $L(C) < \frac{L_۲}{۲}$؟ مقادیری از q ارائه دهید که برای هر n داشته باشیم $L(C) = \frac{L_n}{n}$.

فصل ۳

آنتروپی

اختصار روح عقل است. (Hamlet)

هدف این فصل معرفی تابع آنتروپی است که مقدار اطلاعات خارج‌شده از یک منبع را اندازه‌گیری می‌کند. خواص اصلی این تابع را بررسی کرده و چگونگی ارتباط آن‌را با متوسط طول کد کدهای منبع نشان می‌دهیم.

۱.۳ اطلاعات و آنتروپی

برای تعیین مقدار اطلاعات حمل‌شده توسط سمبل‌های s_i از یک منبع $\mathcal{S}$، یک عدد $I(s_i)$ معرفی می‌کنیم که معرف مقدار اطلاعات حاصل از دانستن این است که منبع $\mathcal{S}$ سمبل s_i را تولید کرده است؛ این عدد همچنین معرف مقدار ابهام اولیه[۱] در این‌که آیا s_i اتفاق خواهد افتاد، و تعجب ما بعد از مطلع شدن از این‌که این سمبل ظاهر شده است می‌باشد. از این‌رو لازم است که:

(۱) $I(s_i)$ یک تابع نزولی از احتمال p_i، متناظر با s_i، است با خاصیت
اگر $p_i = ۱$ آنگاه $I(s_i) = ۰$؛

$$I(s_i s_j) = I(s_i) + I(s_j). \qquad (۲)$$

بر اساس خاصیت (۱) هرچه احتمال یک پیشامد بیشتر باشد اطلاعات آزاد شده کمتر بوده، و یک پیشامد غیرقابل اجتناب حامل هیچ اطلاعاتی نیست. روزنامه‌نگاران اصولاً از این

[۱] prior uncertainty

خاصیت در انتخاب موضوع استفاده می‌کنند. شرط (۲) می‌گوید که چون منبع S سمبل‌ها را به‌طور مستقل تولید می‌کند میزان اطلاعات به‌دست آمده از دانستن دو سمبل متوالی برابر است با مجموع دو مقدار اطلاعات متناظر با هر یک از آن دو سمبل. (اگر سمبل‌های متوالی مستقل نباشند این مقدار باید از مجموع کمتر باشد زیرا با دانستن s_i چیزی را در رابطه با s_j به‌دست می‌آوریم.)

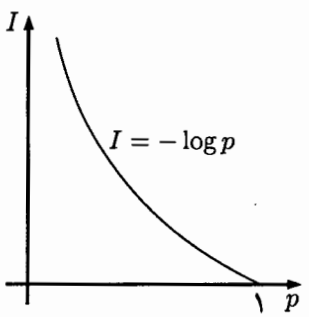
استقلال سمبل‌ها در S به این معنی است که برای هر i و j رابطه $Pr(s_is_j) = Pr(s_i)Pr(s_j) = p_ip_j$ برقرار است. نتیجه می‌شود که شرایط (۱) و (۲) برقرار می‌شوند اگر تعریف کنیم:

$$I(s_i) = -\log p_i = \log \frac{1}{p_i}, \qquad (1.3)$$

و بنابراین

$$I(s_is_j) = \log \frac{1}{p_ip_j} = \log \frac{1}{p_i} + \log \frac{1}{p_j} = I(s_i) + I(s_j).$$

چون اگر $p_i \longrightarrow 0$ آنگاه $I(s_i) \longrightarrow \infty$، از قرارداد $I(s_i) = \infty$ برای $p_i = 0$ استفاده می‌کنیم. گراف این تابع در شکل ۱.۳ نشان داده شده است.



شکل ۱.۳

مبنای انتخاب شده برای لگاریتم خیلی مهم نیست. معمولاً از لگاریتم در پایه r که r تعداد سمبل‌های کد است استفاده شده، و از این‌رو در حالت کدهای دوتایی از $\log = \lg = \log_2$ استفاده می‌کنیم. تغییر مبنای یک لگاریتم صرفاً معرف یک تغییر در واحد است: چون برای هر $x > 0$ داریم

$$x = r^{\log_r x}$$

پس با گرفتن لگاریتم در مبنای s نتیجه می‌شود

$$\log_s x = \log_s r . \log_r x.$$

در حالت دوتایی، واحد اطلاعات را بیت bits (binary digits) می‌نامند. اگر r مهم نبوده و یا دانسته فرض شود می‌نویسیم $I(s_i) = -\log(p_i)$؛ در صورت تمایل به تاکید روی r از $I_r(s_i) = -\log_r(p_i)$ استفاده می‌کنیم.

مثال ۱.۳ فرض کنید S یک سکه نااریب بوده و s_1 و s_2 معرف رو (heads) و پشت (tails) باشند. دراین‌صورت $p_1 = p_2 = \frac{1}{2}$ و بنابراین اگر قرار دهیم $r = 2$ آنگاه $I_2(s_1) = I_2(s_1) = 1$. پس واحد استاندارد اطلاعات برابر مقداری است که از یک‌بار پرتاب یک سکه نااریب حاصل می‌شود.

چون هر سمبل s_i از منبع S با احتمال p_i ظاهر می‌شود نتیجه می‌شود که میانگین اطلاعات حمل‌شده به‌وسیله S با توزیع زیر مشخص می‌شود

$$H_r(S) = \sum_{i=1}^{q} p_i I_r(s_i) = \sum_{i=1}^{q} p_i \log_r \frac{1}{p_i} = -\sum_{i=1}^{q} p_i \log_r p_i,$$

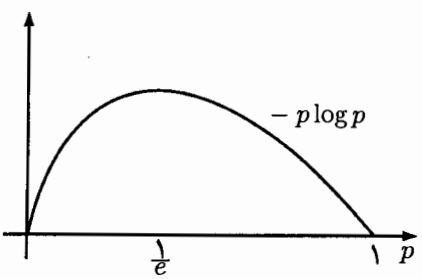
که آنتروپی r–تایی S نامیده می‌شود. همچون تابع I، یک تغییر مبنا در لگاریتم متناظر با یک تغییر واحد است که با رابطه زیر معرفی می‌شود

$$H_s(S) = \log_s r . H_r(S).$$

وقتی‌که r مشخص است و یا مهم نیست، صرفاً می‌نویسیم

$$H(S) = \sum_{i=1}^{q} p_i \log \frac{1}{p_i} = -\sum_{i=1}^{q} p_i \log p_i. \qquad (۲.۳)$$

چون اگر $p \longrightarrow 0$ آنگاه $p\log(\frac{1}{p}) = -p\log p \longrightarrow 0$ (شکل ۲.۳)، قرارداد $p\log\frac{1}{p} = 0$ را برای $p = 0$ به‌کار می‌بریم، و بنابراین $H(S)$ یک تابع پیوسته از احتمال‌های p_i است.



شکل ۲.۳

مثال ۲.۳ فرض کنید S دارای دو سمبل با احتمال‌های p و $p-1$ باشد؛ پس S می‌تواند پرتاب یک سکه احتمالاً اریب باشد. به‌عوض تعداد دفعات وقوع صرفاً این توزیع احتمال را به‌کار برده، برای راحتی نماد $\bar{p} = 1 - p$ را معرفی می‌کنیم که $0 \leq p \leq 1$. (ابهامی در رابطه با مزدوج مختلط که در این کتاب به‌کار نمی‌روند نباید ایجاد شود.) بنابراین

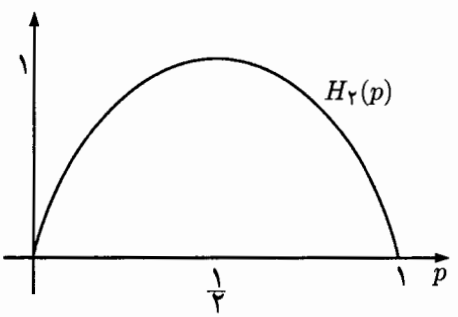
$$H(S) = -p\log p - \bar{p}\log\bar{p}.$$

ما این تابع مهم را با $H(p)$، یا دقیق‌تر $H_r(p)$، نمایش می‌دهیم؛ پس

$$H(p) = -p\log p - \bar{p}\log\bar{p}. \qquad (۳.۳)$$

گراف تابع $H_2(p)$ با شکل ۳.۳ نمایش داده شده است؛ برای حالت کلی r، مقیاس عمودی را در فاکتور $\log_r 2$ ضرب می‌کنیم. این نشان می‌دهد که ماکزیمم و می‌نیمم $H(p)$ برابر ۱ و ۰ بوده که به‌ترتیب متناظر با $p = \frac{1}{2}$ و $p = 0$ و $p = 1$ هستند. بنابراین بیشترین و کمترین عدم اطمینان نسبت به S متناظر با بیشترین و کمترین اطلاعات حمل‌شده به‌وسیله S هستند. توجه کنید که گراف نسبت به خط عمودی $p = \frac{1}{2}$ تقارن دارد، یعنی این‌که

$$H(p) = H(\bar{p}).$$



شکل ۳.۳

اگر در مثال ۲.۳ قرار دهیم $p = \frac{2}{3}$ (مانند بخش ٦.٢)، و از تقریب $\log_2 3 \approx 1.585$ استفاده شود خواهیم دید که

$$H_2(S) = \frac{2}{3}\log_2\frac{3}{2} + \frac{1}{3}\log_2 3 = \log_2 3 - \frac{2}{3}\log_2 2 = \log_2 3 - \frac{2}{3} \approx 0.918.$$

بنابراین این سکه اریب نسبتاً اطلاعات کمتری را در مقایسه با سکه درنظر گرفته شده در مثال ۱.۳، که $H_2(S) = 1$، در بر دارد.

مثال ۳.۳ اگر S دارای پنج سمبل با احتمال‌های $p_i = ۰.۳, ۰.۲, ۰.۲, ۰.۲, ۰.۱$ (مانند بخش ۲.۲، مثال ۵.۲) باشد ملاحظه می‌کنیم که $H_۲(S) \approx ۲.۲۴۶$.

مثال ۴.۳ اگر S دارای q سمبل با احتمال یکسان باشد آنگاه برای هر i داریم $p_i = \frac{۱}{q}$، و از این‌رو

$$H_r(S) = q.\frac{۱}{q}\log_r q = \log_r q.$$

بالاخص، اگر قرار دهیم $q = ۵$ (مانند مثال ۶.۲)، آنگاه $H_۲(S) = \log_۲ ۵ \approx ۲.۳۲۱$. با مقایسه این با مثال ۳.۳، می‌بینیم که چگونه یک منبع با توزیع یکنواخت اطلاعات بیشتری را در مقایسه با یک منبع با احتمال‌های متنوع در بر دارد.

آنتروپی این منابع را همچنین می‌توان با میانگین طول کدهای به‌دست آمده به‌وسیله کدهای هافمن دوتایی فصل ۲ مقایسه نمود. در مثال ۲.۳، به‌عنوان نمونه، می‌بینیم که با $p = \frac{۲}{۳}$ و $n = ۱, ۲, ۳$ متوسط طول کد به‌دست آمده به‌وسیله کد دوتایی هافمن برای S^n در بخش ۶.۲ به‌ترتیب برابر $L \approx ۱, ۰.۹۴۴, ۰.۹۳۸$ است که به آنتروپی $H_۲(S) \approx ۰.۹۱۸$ نزدیک می‌شوند. در مثال ۳.۳، متوسط طول کد $L(C) = ۲.۳$ که در مثال ۵.۲ حاصل شد نزدیک به آنتروپی $H_۲(S) \approx ۲.۲۴۶$ است. مشابهاً در مثال ۴.۳، متوسط طول کد $L(C) = ۲.۴$ که در مثال ۶.۲ به‌دست آمده نزدیک به آنتروپی $H_۲(S) \approx ۲.۳۲۱$ است. این ارتباط نزدیک بین آنتروپی و میانگین طول کد توضیح دهنده قضیه اول شانون[۱] است که در بخش ۶.۳ بیان و ثابت خواهد شد.

مثال ۵.۳ با قرار دادن $q = ۶$ در مثال ۴.۳، می‌بینیم که یک تاس نااریب دارای آنتروپی $\log_۲ ۶ \approx ۲.۵۸۶$ است.

مثال ۶.۳ با به‌کار بردن فراوانی معلوم حروف الفبا، آنتروپی متن انگلیسی تقریباً برابر ۴.۰۳ محاسبه شده است.

این مثال آخر، که ظاهراً حکایت از آن دارد که خواندن یک کتاب حدود چهار برابر پرتاب یک سکه اطلاعات در بر دارد، توضیح می‌دهد که در واقع نظریه اطلاعات مربوط به میزان مفید بودن و یا جالب بودن یک پیام نمی‌شود، زیرا این‌ها بستگی خیلی زیاد به شخصی دارند که آن پیام را می‌خواند. یک آماردان ممکن است از دریافت یک کتاب درباره اعداد یا حروف تصادفی خوشحال شود، درحالی‌که یک شخص عادی احتمالاً یک کتاب داستان را ترجیح می‌دهد حتی اگر از آنتروپی کمتری برخوردار باشد.

---

[۱] Shannon's first theorem

تمرین ۱.۳ یک منبع $\mathcal{S}$ دارای توزیع احتمال

$$p_i = ۰.۳, ۰.۲, ۰.۱۵, ۰.۱, ۰.۱, ۰.۰۸, ۰.۰۵, ۰.۰۲$$

است. مقادیر $H_۲(\mathcal{S})$ و $H_۳(\mathcal{S})$ را پیدا کرده و این مقادیر را با متوسط طول کلمه برای کدهای هافمن دوتایی و سه‌تایی منبع $\mathcal{S}$ مقایسه کنید (تمرین ۷.۲ را ببینید).

۲.۳ خواص تابع آنتروپی

در بخش ۱.۳ آنتروپی یک منبع $\mathcal{S}$ با احتمال‌های p_i را برابر $H_r(\mathcal{S})$ تعریف کردیم:

$$H_r(\mathcal{S}) = \sum_i p_i \log_r \frac{۱}{p_i}.$$

چون $p \log_r(\frac{۱}{p}) \geq ۰$ و تساوی برقرار است اگر و فقط اگر $p = ۰$ یا $p = ۱$، قضیه زیر را داریم.

قضیه ۷.۳ $H_r(\mathcal{S}) \geq ۰$ و تساوی برقرار است اگر و فقط اگر $p_i = ۰$ به‌جز یکی از آن‌ها که برابر ۱ می‌باشد.

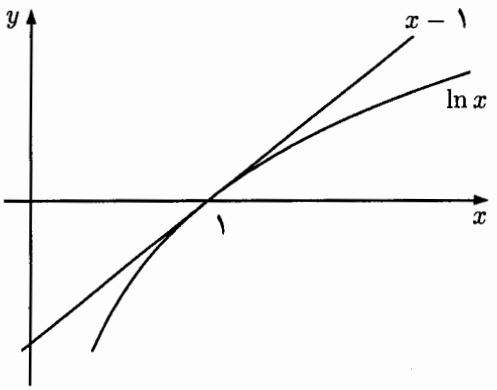
بنابراین کمترین مقدار آنتروپی زمانی است که هیچ ابهامی در رابطه با سمبل‌های ظاهر شده از $\mathcal{S}$ وجود نداشته باشد، یعنی همیشه یک سمبل است که ظاهر می‌شود، و از این‌رو هیچ اطلاعاتی آزاد نمی‌شود. چه موقع آنتروپی ماکزیمم می‌شود؟ برای جواب به این سؤال به لم زیر نیاز داریم.

لم ۸.۳ برای هر $x > ۰$ رابطه $\ln x \leq x - ۱$ برقرار بوده و تساوی رخ می‌دهد اگر و فقط اگر $x = ۱$.

اثبات قرار دهید $f(x) = x - ۱ - \ln x$، پس $f(۱) = ۰$. سپس برای هر $x > ۰$ داریم $f'(x) = ۱ - x^{-۱}$، از این‌رو f دارای تنها نقطه اکسترمم در $x = ۱$ است. چون برای هر x داریم $f''(x) = x^{-۲} > ۰$ این نقطه تنها می‌نیمم مطلق f بوده و بنابراین $f \geq ۰$، و تساوی برقرار است اگر و فقط اگر $x = ۱$. ■

این نتیجه در شکل ۴.۳ تشریح شده است.

با تغییر مبنای لگاریتم به r، نتیجه می‌شود $\log_r x \leq (x - ۱) \log_r e$، و تساوی برقرار است اگر و فقط اگر $x = ۱$. نتیجه بعدی نسبتاً فنی به‌نظر می‌رسد، ولی حاوی چند نتیجه خیلی مفید است.



شکل ۴.۳

نتیجه ۹.۳ فرض کنید $x_i \geq ۰$ و $y_i > ۰$ که $۰ \leq i \leq q$، و فرض کنید $\sum_i x_i = ۱$ و $\sum_i y_i = ۱$ (پس (x_i) و (y_i) توزیع احتمال بوده با این خاصیت که $y_i \neq ۰$). در این‌صورت

$$\sum_{i=۱}^{q} x_i \log_r \frac{۱}{x_i} \leq \sum_{i=۱}^{q} x_i \log_r \frac{۱}{y_i},$$

یعنی این‌که $\sum_i x_i \log_r(y_i/x_i) \leq ۰$، و تساوی برقرار است اگر و فقط اگر برای هر i داشته باشیم $x_i = y_i$.

اثبات اگر تمام x_i ها ناصفر باشند آنگاه تفاضل بین سمت چپ و راست نامساوی را درنظر بگیرید:

$$\begin{aligned}
LHS - RHS &= \sum_{i=۱}^{q} x_i \log_r(\frac{y_i}{x_i}) \\
&= \frac{۱}{\ln r} \sum_{i=۱}^{q} x_i \ln(\frac{y_i}{x_i}) && (\log_r x = \tfrac{\ln x}{\ln r} \text{ چون}) \\
&\leq \frac{۱}{\ln r} \sum_{i=۱}^{q} x_i(\frac{y_i}{x_i} - ۱) && (q \text{ با راستفاده از لم } ۸.۳) \\
&= \frac{۱}{\ln r}(\sum_{i=۱}^{q} y_i - \sum_{i=۱}^{q} x_i) \\
&= ۰,
\end{aligned}$$

که در این‌جا تساوی برقرار است اگر و فقط اگر همواره $y_i = x_i$. وقتی بعضی از x_i ها ۰ باشند بحث مشابه‌ای برقرار است زیرا با توجه به قرارداد $x_i \log(۱/x_i) = ۰$ می‌توان یک چنین جمله‌ای را ندیده گرفت. ■

قضیه ۱۰.۳ اگر یک منبع $\mathcal{S}$ دارای q سمبل باشد آنگاه $H_r(\mathcal{S}) \leq \log_r q$، و تساوی برقرار است اگر و فقط اگر منبع دارای توزیع احتمال یکنواخت باشد.

اثبات اگر قرار دهیم $x_i = p_i$ (احتمال‌های منبع $\mathcal{S}$) و $y_i = \frac{1}{q}$، آنگاه شرایط لم ۹.۳ برقرار می‌شوند. بنابراین داریم

$$H_r(\mathcal{S}) = \sum_{i=1}^{q} p_i \log_r \frac{1}{p_i} \leq \sum_{i=1}^{q} p_i \log_r q = \log_r q \sum_{i=1}^{q} p_i \log_r q,$$

و تساوی برقرار است اگر و فقط اگر $p_i = \frac{1}{q}$. ■

پس زمانی که بیشترین ابهام در مورد سمبل‌های ظاهر شده وجود داشته باشد آنتروپی ماکزیمم بوده، و بیشترین مقدار اطلاعات حمل می‌شود.

۳.۳ آنتروپی و متوسط طول کلمه

در اواخر بخش ۱.۳، چند منبع را درنظر گرفته و آنتروپی آنها را با متوسط طول کلمه کدهای دوتایی هافمن آنها مقایسه کردیم. اکنون رابطه بین آنتروپی و متوسط طول کد را با جزییات بیشتر بررسی می‌کنیم.

قضیه ۱۱.۳ اگر $\mathcal{C}$ یک کد یکتا دکدپذیر r-تایی برای منبع $\mathcal{S}$ باشد، آنگاه

$$L(\mathcal{C}) \geq H_r(\mathcal{S}).$$

اثبات اگر تعریف کنیم

$$K = \sum_{i=1}^{q} r^{-l_i},$$

که در آن $l_1, \ldots, l_q$ طول کلمات $\mathcal{C}$ هستند، آنگاه از نامساوی مک‌میلان (قضیه ۲۱.۱) نتیجه می‌شود $K \leq 1$. حال نتیجه ۹.۳ را با $x_i = p_i$ و $y_i = r^{-l_i}/K$ به‌کار می‌بریم، پس $y_i > 0$ و $\sum_i y_i = 1$. در این‌صورت

$$\begin{aligned}
H_r(\mathcal{S}) &= \sum_{i=1}^{q} p_i \log_r(\frac{1}{p_i}) \\
&\leq \sum_{i=1}^{q} p_i \log_r(\frac{1}{y_i}) \qquad \text{(بنابر نتیجه ۹.۳)} \\
&= \sum_{i=1}^{q} p_i \log_r(r^{l_i} K) \\
&= \sum_{i=1}^{q} p_i (l_i + \log_r K)
\end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^{q} p_i l_i + \log_r K \sum_{i=1}^{q} p_i \\ &= L(\mathcal{C}) + \log_r K \qquad (\text{چون } \sum p_i = 1) \\ &\leq L(\mathcal{C}) \end{aligned}$$

(چون از $K \leq 1$ نتیجه می‌شود $\log_r K \leq 0$). ■

تفسیر این موضوع چنین است: هر سمبل ظاهر شده از $\mathcal{S}$ به‌طور متوسط $H_r(\mathcal{S})$ واحد از اطلاعات حمل می‌کند؛ اگر $\mathcal{S}$ را بخواهیم بدون از دست دادن این اطلاعات کدگذاری کنیم آنگاه کد $\mathcal{C}$ باید یکتا دکدپذیر باشد؛ هر سمبل از کد یک بیت از اطلاعات حمل می‌کند، بنابراین به‌طور متوسط هر کدکلمه از $\mathcal{C}$ باید حاوی حداقل $H_r(\mathcal{S})$ کدسمبل باشد، یعنی این‌که $L(\mathcal{C}) \geq H_r(\mathcal{S})$. بالاخص منابعی که اطلاعات بیشتری آزاد می‌کنند کدکلمه‌های طولانی‌تری نیاز دارند.

نتیجه ۱۲.۳ برای یک منبع مفروض $\mathcal{S}$ با احتمال‌های p_i، یک کد یکتا دکدپذیر $\mathcal{C}$ با خاصیت $L(\mathcal{C}) = H_r(\mathcal{S})$ وجود دارد اگر و فقط اگر برای هر i عدد $\log_r p_i$ صحیح باشد، یعنی این‌که برای هر i رابطه $p_i = r^{e_i}$ به‌ازای عدد صحیحی چون $e_i \leq 0$ برقرار باشد.

اثبات ($\Longrightarrow$) اگر $L(\mathcal{C}) = H_r(\mathcal{S})$، آنگاه هر دو نامساوی موجود در اثبات قضیه ۱۱.۳ باید معرف تساوی باشند. پس بنابه نتیجه ۹.۳ به‌ازای هر i داریم $p_i = y_i$ و $\log_r K = 0$. پس $K = 1$ و $p_i = r^{-l_i}/K = r^{-l_i}$، و بنابراین $\log_r p_i = -l_i$.

($\Longleftarrow$) فرض کنید که برای هر i عدد $-\log_r p_i$ – یک عدد صحیح l_i باشد. چون $p_i \leq 1$ پس $l_i \geq 0$. اکنون برای هر i داریم $r^{l_i} = 1/p_i$ و از این‌رو

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} = \sum_{i=1}^{q} p_i = 1.$$

بنابراین نامساوی مک‌میلان (قضیه ۲۱.۱) برقرار بوده، و از این‌رو یک کد r–تایی یکتا دکدپذیر $\mathcal{C}$ برای $\mathcal{S}$ با طول کلمات l_i وجود دارد. این کد دارای میانگین طول کلمه $H_r(\mathcal{S})$ است

$$L(\mathcal{C}) = \sum_{i=1}^{q} p_i l_i = \sum_{i=1}^{q} p_i \log_r \frac{1}{p_i} = H_r(\mathcal{S}). \quad \blacksquare$$

شرط $p_i = r^{e_i}$ در نتیجه ۱۲.۳ خیلی محدود کننده است؛ برای اکثر منابع، هر کد یکتا دکدپذیر دارای خاصیت $L(\mathcal{C}) > H_r(\mathcal{S})$ است.

مثال ۱۳.۳ اگر منبع $\mathcal{S}$ دارای $q = ۳$ سمبل s_i با احتمال‌های $p_i = \frac{۱}{۴}, \frac{۱}{۲}, \frac{۱}{۴}$ (مثال‌های ۲.۱ و ۱.۲) باشد، آنگاه آنتروپی دوتایی $\mathcal{S}$ برابر است با

$$H_۲(\mathcal{S}) = \frac{۱}{۴}\log_۲ ۴ + \frac{۱}{۲}\log_۲ ۲ + \frac{۱}{۴}\log_۲ ۴ = \frac{۱}{۴}.۲ + \frac{۱}{۲}.۱ + \frac{۱}{۴}.۲ = \frac{۳}{۲}.$$

کد $\mathcal{C}$ با ضابطه $s_۱ \longmapsto ۰۰$، $s_۲ \longmapsto ۱$، $s_۳ \longmapsto ۰۱$ یک کد هافمن دوتایی برای $\mathcal{S}$ بوده و از این‌رو بهینه است. این کد دارای متوسط طول کلمه $\frac{۳}{۲}$ است

$$L(\mathcal{C}) = \frac{۱}{۴}.۲ + \frac{۱}{۲}.۱ + \frac{۱}{۴}.۲ = \frac{۳}{۲}.$$

پس در این حالت $L(\mathcal{C}) = H_۲(\mathcal{S})$. علت این وضع این است که احتمال‌های p_i توان‌های صحیحی از ۲ هستند.

مثال ۱۴.۳ همچون مثال ۵.۲ از بخش ۲.۲، فرض کنید $\mathcal{S}$ دارای $q = ۵$ سمبل با احتمال‌های $p_i = ۰.۳, ۰.۲, ۰.۲, ۰.۲, ۰.۱$ باشد. در مثال ۳.۳ ملاحظه شد که $H_۲(\mathcal{S}) \approx ۲.۲۴۶$، و در مثال ۵.۲ دیدیم که یک کد دوتایی هافمن برای این منبع دارای متوسط طول کلمه ۲.۳ است. پس از قضیه ۸.۲ نتیجه می‌شود که هر کد دوتایی یکتا دکدپذیر $\mathcal{C}$ برای $\mathcal{S}$ در شرط $L(\mathcal{C}) \geq ۲.۳ > H_۲(\mathcal{S})$ صدق می‌کند. از این‌رو هیچ‌یک از آن کدها واجد خاصیت $L(\mathcal{C}) = H_۲(\mathcal{S})$ نیست، و دلیل آن در این حالت این است که همگی احتمال‌های p_i توانی از ۲ نیستند.

تمرین ۲.۳ برای هر $q \geq ۲$، مثالی از یک منبع $\mathcal{S}$ با q سمبل و یک کد لحظه‌ای دوتایی $\mathcal{C}$ برای $\mathcal{S}$ ارائه دهید که در شرط $L(\mathcal{C}) = H_۲(\mathcal{S})$ صدق کند.

بنابر نتیجه ۱۲.۳، اگر بعضی از احتمال‌های p_i برابر ۰ باشند آنگاه باید داشته باشیم $L(\mathcal{C}) > H_۲(\mathcal{S})$. با این حال، با حذف یک چنین سمبل‌هایی ممکن است که قادر به رسیدن به تساوی، با کاهش دادن $L(\mathcal{C})$ به‌وسیله اختصاص کدکلمه‌های کوتاه‌تر و بدون تغییر در آنتروپی، باشیم.

مثال ۱۵.۳ فرض کنید $\mathcal{S}$ دارای سه سمبل با احتمال‌های $p_i = \frac{۱}{۲}, \frac{۱}{۲}, ۰$ باشد. در این‌صورت $H(\mathcal{S}) = ۱$، ولی یک کد دوتایی هافمن $\mathcal{C}$ برای $\mathcal{S}$ دارای طول کدهای ۱ ، ۲ ، ۲ و متوسط طول $L(\mathcal{C}) = ۱.۵$ است. بدون درنظر گرفتن $s_۳$ رابطه $H_۲(\mathcal{S}) = ۱ = L(\mathcal{C})$ برقرار است، زیرا در این حالت می‌توان کد $\mathcal{C} = \{۰, ۱\}$ را به‌کار برد؛ بعد از حذف $s_۳$ تساوی برقرار است زیرا بقیه احتمال‌ها همگی توان صحیح از $r = ۲$ هستند.

اگر C یک کد r–تایی برای یک منبع S باشد کارایی آن را چنین تعریف می‌کنیم:

$$\eta = \frac{H_r(S)}{L(C)}. \qquad (۴.۳)$$

بنابراین از قضیه ۱۱.۳ نتیجه می‌شود که برای هر کد یکتا دکدپذیر C رابطه $۰ \leq \eta \leq ۱$ برقرار است. افزونگی C را برابر $\bar{\eta} = ۱ - \eta$ تعریف می‌کنیم؛ پس افزایش افزونگی باعث کاهش کارایی می‌شود که در تناقض با اعتقاد بعضی از استخدام کنندگان است. در مثال‌های ۱۳.۳ و ۱۴.۳ مقدار η به‌ترتیب برابر $\eta = ۱$ و $\eta \approx ۰.۹۹۷$ است.

تمرین ۳.۳ یک منبع S دارای احتمال‌های ۰.۴، ۰.۳، ۰.۱، ۰.۱، ۰.۰۶، ۰.۰۴ است (تمرین ۳.۲). آنتروپی S را محاسبه کرده و از آن‌جا کارایی یک کد دوتایی هافمن برای S را پیدا کنید.

۴.۳ کدگذاری شانون — فانو

کدهای هافمن بهینه هستند، ولی محاسبه متوسط طول کلمات آن می‌تواند کسل‌کننده باشد. کدهای شانون — فانو[۱] نزدیک به بهینه هستند ولی متوسط طول کلمات آنها ساده‌تر برآورد می‌شود.

اجازه دهید فرض کنیم که احتمال‌های منبع S ناصفر هستند. بنابر نتیجه ۱۲.۳ اگر متوسط طول کد، $L(C)$، از یک کد یکتا دکدپذیر r– تایی C برای منبع S بخواهد برابر $H_r(S)$ باشد آنگاه طول کلمات آن باید در شرط $l_i = \log_r(۱/p_i)$ صدق کنند. این معمولاً غیر ممکن است زیرا اعداد $\log_r(۱/p_i)$ معمولاً صحیح نیستند. در این حالت بهترین انتخاب بعدی را انجام داده و قرار می‌دهیم

$$l_i = \lceil \log_r(۱/p_i) \rceil \qquad (۵.۳)$$

که $\lceil x \rceil$ معرف کوچک‌ترین عدد صحیح بزرگ‌تر یا مساوی x است. (این در واقع تابع سقف است که به عدد صحیح بعدی گرد می‌کند.) پس l_i عدد صحیح منحصربه‌فردی است که دارای خاصیت زیر می‌باشد

$$\log_r \frac{۱}{p_i} \leq l_i < ۱ + \log_r \frac{۱}{p_i}, \qquad (۶.۳)$$

و بنابراین برای هر i داریم $p_i \geq r^{-l_i}$. با جمع زدن روی i داریم

$$K = \sum_{i=۱}^{q} r^{-l_i} \leq \sum_{i=۱}^{q} p_i = ۱.$$

---

[۱] Shannon - Fano coding

از این‌رو قضیه ۲۰.۱ (نامساوی کرفت) ایجاب می‌کند که یک کد r–تایی لحظه‌ای $\mathcal{C}$ با طول کلمات l_i برای $\mathcal{S}$ وجود دارد. کد $\mathcal{C}$ را یک کد شانون – فانو برای $\mathcal{S}$ می‌نامیم. توجه کنید که ما تنها وجود این کدها را ثابت کرده و نحوه ساخت آنها را بررسی نکرده‌ایم.

اگر رابطه (۶.۳) را در p_i ضرب کرده و سپس روی i عمل جمع انجام دهیم خواهیم داشت

$$\sum_{i=1}^{q} p_i \log_r \frac{1}{p_i} \leq \sum_{i=1}^{q} p_i l_i < \sum_{i=1}^{q} p_i (1 + \log_r \frac{1}{p_i}) = 1 + \sum_{i=1}^{q} p_i \log_r \frac{1}{p_i},$$

و از این‌رو

$$H_r(\mathcal{S}) \leq L(\mathcal{C}) < 1 + H_r(\mathcal{S}). \qquad (۷.۳)$$

این بحث را می‌توان به حالتی که بعضی از p_i ها صفر هستند تعمیم داد که این عمل با حد گرفتن، وقتی $p_i \longrightarrow 0$، انجام می‌شود (جزییات را حذف می‌کنیم). ولی حد گرفتن تندی نامساوی‌ها را از بین می‌برد، و بنابراین در این حالت نتیجه اندکی ضعیف‌تر را داریم

$$H_r(\mathcal{S}) \leq L(\mathcal{C}) \leq 1 + H_r(\mathcal{S}). \qquad (۸.۳)$$

بنابراین قضیه زیر را ثابت کرده‌ایم.

قضیه ۱۶.۳ هر کد r–تایی شانون – فانو $\mathcal{C}$ برای یک منبع $\mathcal{S}$ دارای خاصیت زیر است

$$H_r(\mathcal{S}) \leq L(\mathcal{C}) \leq 1 + H_r(\mathcal{S}).$$

نتیجه ۱۷.۳ هر کد r–تایی بهینه $\mathcal{D}$ برای یک منبع $\mathcal{S}$ در شرط زیر صدق می‌کند:

$$H_r(\mathcal{S}) \leq L(\mathcal{D}) \leq 1 + H_r(\mathcal{S}).$$

اثبات از قضایای ۱۱.۳ و ۱۶.۳ نتیجه می‌شود

$$H_r(\mathcal{S}) \leq L(\mathcal{D}) \leq L(\mathcal{C}) \leq 1 + H_r(\mathcal{S}). \quad \blacksquare$$

این بدین معنی است که حتی اگر به $H_r(\mathcal{S})$ نتوان رسید می‌توان کدهایی پیدا کرد که به‌طور معقولی به آن نزدیک باشند.

مثال ۱۸.۳ فرض کنید همچون مثال ۵.۲ منبع $\mathcal{S}$ دارای ۵ سمبل با احتمال‌های ۰.۱، ۰.۲، ۰.۲، ۰.۲، ۰.۳ باشد. پس $1/p_i = 10/3, 5, 5, 5, 10$. بنابراین یک کد دوتایی شانون – فانو $\mathcal{C}$ برای $\mathcal{S}$ دارای طول کلمه l_i است

$$l_i = \lceil \log_2(1/p_i) \rceil = \min\{n \in Z \mid 2^n \geq 1/p_i\} = 2, 3, 3, 3, 4;$$

و در نتیجه متوسط طول کلمه آن برابر $L(\mathcal{C}) = \sum p_i l_i = ۲.۸$ است. این را با یک کد هافمن $\mathcal{D}$ برای $\mathcal{S}$ مقایسه کنید که دارای متوسط طول $L(\mathcal{D}) = ۲.۳$ است. در مثال ۳.۳ دیدیم که $H_۲(\mathcal{S}) \approx ۲.۲۴۶$، پس $\mathcal{C}$ در شرایط قضیه ۱۶.۳ صدق می‌کند. کارایی $\mathcal{C}$ برابر $\eta \approx ۲.۲۴۶/۲.۸ \approx ۰.۸۰۲$ است حال آن‌که $\mathcal{D}$ کارایی $۰.۹۷۷ \approx ۲.۲۴۶/۲.۳ \approx \eta$ دارد.

مثال ۱۹.۳ اگر $p_۱ = ۱$ و $p_i = ۰$ برای $i > ۱$، آنگاه $H_r(\mathcal{S}) = ۰$. یک کد r–تایی بهینه $\mathcal{D}$ برای $\mathcal{S}$ دارای میانگین طول کلمه $L(\mathcal{D}) = ۱$ است، پس در این حالت کران بالای $۱ + H_r(\mathcal{S})$ حاصل شده است.

تمرین ۴.۳ یک منبع $\mathcal{S}$ با احتمال‌های ۰.۰۴، ۰.۰۶، ۰.۱، ۰.۱، ۰.۳، ۰.۴ مفروض است. در رابطه با یک کد دوتایی شانون — فانو $\mathcal{D}$ برای این منبع طول کلمات، متوسط طول، و کارایی آن‌را محاسبه کنید. نتیجه را با تمرین ۳.۳ که مربوط به کد بهینه برای $\mathcal{S}$ است مقایسه کنید.

در حالت کلی کدهای شانون — فانو از بهینگی دور نیستند. اگر از آنها برای کدکردن بسط یک منبع استفاده شود به بهینگی نزدیک‌تر می‌شوند. به‌منظور کسب آمادگی برای اثبات این نتیجه در بخش ۶.۳، آنتروپی بسط منابع را در بخش بعد بررسی می‌کنیم.

۵.۳ آنتروپی بسط‌ها و ضرب‌ها

از بخش ۶.۲ به‌خاطر بیاورید که منبع $\mathcal{S}^n$ دارای q^n سمبل $s_{i_۱} \dots s_{i_n}$ با احتمال‌های $p_{i_۱} \dots p_{i_n}$ است. اگر $\mathcal{S}^n$ را به‌عنوان n کپی مستقل از $\mathcal{S}$ درنظر بگیریم آنگاه باید انتظار این‌را داشته باشیم که اطلاعات برآمده از آن n برابر اطلاعات مربوط به $\mathcal{S}$ باشد. این قضیه بعد را القا می‌کند.

قضیه ۲۰.۳ اگر $\mathcal{S}$ یک منبع دلخواه باشد آنگاه $H_r(\mathcal{S}^n) = nH_r(\mathcal{S})$.

قبل از اثبات این قضیه، باید ابتدا تعریف بسط را با درنظر گرفتن ضرب منابع تعمیم دهیم. فرض کنید $\mathcal{S}$ و $\mathcal{T}$ دو منبع با سمبل‌های s_i و t_j و با احتمال‌های p_i و q_j باشند. ضرب $\mathcal{S} \times \mathcal{T}$ را برابر منبعی تعریف می‌کنیم که سمبل‌های آن زوج‌های (s_i, t_j)، مختصراً $s_i t_j$، با احتمال‌های $Pr(s_i, t_j)$ باشند. می‌توان $\mathcal{S} \times \mathcal{T}$ را به‌عنوان یک زوج متشکل از $\mathcal{S}$ و $\mathcal{T}$ درنظر گرفت که به‌طور هم‌زمان سمبل‌های s_i و t_j را تولید می‌کند. گوییم $\mathcal{S}$ و $\mathcal{T}$ مستقل هستند اگر همواره $Pr(s_i, t_j) = p_i q_j$. به‌عنوان نمونه $\mathcal{S}$ و $\mathcal{T}$ می‌توانند معرف هوای دو شهر دور از یکدیگر باشند (ولی نه نزدیک به هم، زیرا در آن‌صورت دیگر مستقل از هم نخواهند بود). بسط $\mathcal{S}^۲$ را می‌توان به‌عنوان ضرب $\mathcal{S} \times \mathcal{S}$ از دو کپی مستقل از هم از منبع $\mathcal{S}$ درنظر گرفت: یک زوج یکسان ولی مستقل از هم تاس مثال خوبی در این رابطه است.

لم ۲۱.۳ اگر $\mathcal{S}$ و $\mathcal{T}$ منابع مستقلی باشند آنگاه $H_r(\mathcal{S}\times\mathcal{T}) = H_r(\mathcal{S}) + H_r(\mathcal{T})$.

اثبات از مستقل بودن نتیجه می‌شود $Pr(s_it_j) = p_iq_j$، پس با توجه به $\sum p_i = \sum q_j = ۱$ داریم:

$$\begin{aligned}
H_r(\mathcal{S}\times\mathcal{T}) &= -\sum_i\sum_j p_iq_j \log_r p_iq_j \\
&= -\sum_i\sum_j p_iq_j(\log_r p_i + \log_r q_j) \\
&= -\sum_i\sum_j p_iq_j \log_r p_i - \sum_i\sum_j p_iq_j \log_r q_j \\
&= \left(-\sum_i p_i \log_r p_i\right)\left(\sum_j q_j\right) + \left(\sum_i p_i\right)\left(-\sum_j q_j \log_r q_j\right) \\
&= H_r(\mathcal{S}) + H_r(\mathcal{T}). \ \blacksquare
\end{aligned}$$

با به‌کار بردن استقرا می‌توان ضرب هر تعداد متناهی از منابع را تعریف کرد: برای این منظور تعریف می‌کنیم

$$\mathcal{S}_۱ \times \cdots \times \mathcal{S}_n = (\mathcal{S}_۱ \times \cdots \times \mathcal{S}_{n-۱}) \times \mathcal{S}_n.$$

منابع $\mathcal{S}_i$ مستقل هستند هرگاه هر سمبل $s_{i_۱}\dots s_{i_n}$ دارای احتمال $p_{i_۱}\dots p_{i_n}$ باشد که در آن هر s_{i_j} با احتمال p_{i_j} است.

نتیجه ۲۲.۳ اگر $\mathcal{S}_۱$،...،$\mathcal{S}_n$ منابع مستقلی باشند آنگاه

$$H_r(\mathcal{S}_۱ \times \cdots \times \mathcal{S}_n) = H_r(\mathcal{S}_۱) + \cdots + H_r(\mathcal{S}_n).$$

اثبات حکم با استقرا روی n و استفاده از لم ۲۱.۳ ثابت می‌شود.■

اگر $\mathcal{S}_۱$،...،$\mathcal{S}_n$ نمونه‌های مستقلی از منبع $\mathcal{S}$ باشند آنگاه $\mathcal{S}_۱ \times \cdots \times \mathcal{S}_n = \mathcal{S}^n$، و قضیه ۲۰.۳ فوراً از نتیجه ۲۲.۳ به‌دست می‌آید.

۶.۳ قضیه اول شانون

قضیه ۱۱.۳ اظهار می‌دارد که هر کد r–تایی یکتا دکدپذیر C برای منبع $\mathcal{S}$ دارای متوسط طول کلمه $L(C) \geq H_r(\mathcal{S})$ است، و نتیجه ۱۲.۳ ایجاب می‌کند که این کران پایین معمولاً حاصل نمی‌شود. با این حال، نشان خواهیم داد که ایده معرفی شده در پایان بخش ۶.۲، مبنی بر به‌کار بردن یک کد بهینه برای $\mathcal{S}^n$ به‌عنوان کدی برای $\mathcal{S}$، این امکان را می‌دهد که

$\mathcal{S}$ را به‌گونه‌ای کد کنیم که متوسط طول کلمه آن به‌قدر دلخواه به $H_r(\mathcal{S})$ نزدیک باشد وقتی n به‌قدر کافی بزرگ اختیار شود.

به‌خاطر بیاورید که اگر یک کد برای $\mathcal{S}^n$ دارای متوسط طول L_n باشد، آنگاه به‌عنوان یک کد برای $\mathcal{S}$ دارای متوسط طول L_n/n است. بنابر نتیجه ۱۷.۳، یک کد r–تایی بهینه برای $\mathcal{S}^n$ دارای متوسط طول کلمه L_n است که واجد خاصیت زیر می‌باشد

$$H_r(\mathcal{S}^n) \leq L_n \leq ۱ + H_r(\mathcal{S}^n),$$

پس بنابر قضیه ۲۰.۳

$$nH_r(\mathcal{S}) \leq L_n \leq ۱ + nH_r(\mathcal{S}).$$

با تقسیم کردن بر n نتیجه می‌شود

$$H_r(\mathcal{S}) \leq \frac{L_n}{n} \leq \frac{۱}{n} + H_r(\mathcal{S}),$$

و بنابراین

$$\lim_{n\to\infty} \frac{L_n}{n} = H_r(\mathcal{S}).$$

بدین طریق قضیه اول شانون، یا قضیه کدگذاری بدون پارازیت، اثبات می‌شود که به‌وسیله شانون در مقاله اساسی او [Sh 48] چاپ شد. بیان کامل آن به شرح زیر است.

قضیه ۲۳.۳ با انتخاب n به‌قدر کافی بزرگ و کدکردن $\mathcal{S}^n$، می‌توان یک کدگذاری r–تایی یکتا دکدپذیر برای منبع $\mathcal{S}$ ارائه داد که متوسط طول کلمه آن به‌قدر کافی به آنتروپی $H_r(\mathcal{S})$ نزدیک باشد.

در بخش ۶.۲ یک مثال ساده از این نوع را برای $n = ۱, ۲, ۳$ درنظر گرفتیم؛ در بخش بعد با یک n دلخواه به این مثال باز خواهیم گشت.

هزینه به‌کار بردن این قضیه این است که چون در بسیاری از حالات سرعت میل کردن L_n/n به $H_r(\mathcal{S})$ نسبتاً کند است ممکن است که برای داشتن یک کدگذاری کارا به یک مقدار کاملاً بزرگی از n نیاز باشد. حال اگر $\mathcal{S}$ دارای q سمبل باشد آنگاه $\mathcal{S}^n$ دارای q^n سمبل خواهد بود، عددی که با افزایش n به‌سرعت بزرگ می‌شود. این بدین معنی است که ساخت کد و روند کدگذاری کردن هر دو پیچیده و زمان‌بر هستند. عمل کدگشایی نیز به دلیل منتظر ماندن برای دریافت کامل یک قالب از n سمبل با تاخیر صورت می‌گیرد. از این‌رو ممکن است مجبور شویم که با یک مقدار کوچک‌تری از n کار کنیم.

۷.۳ مثالی برای قضیه اول شانون

همچون مثال ۲.۳، فرض کنید S یک منبع با دو سمبل s_1 و s_2 با احتمال‌های $p_1 = \frac{2}{3}$ و $p_2 = \frac{1}{3}$ باشد. در بخش ۱.۳ دیدیم که $H_2(S) = \log_2 3 - \frac{2}{3} \approx 0.918$. نیز در بخش ۶.۲ با به‌کار بردن کد دوتایی هافمن برای S^n، با $n = 1, 2, 3$، متوسط طول کلمه‌های L_n/n را به‌دست آوردیم. برای n های بزرگ‌تر استفاده از کدهای شانون — فانو به‌جای کدهای هافمن راحت‌تر است؛ اگرچه این کدها از کارایی قدری ضعیف‌تر برخوردار هستند ولی کار کردن با آنها آسان‌تر بوده و با میل کردن n به بینهایت عدد L_n/n به $H_r(S)$ میل می‌کند.

تعداد 2^n سمبل s در S^n وجود دارد که هر یک از آنها قالب‌های n تایی از سمبل‌های s_1 یا s_2 می‌باشند. اگر در سمبل s تعداد k سمبل s_1 (و بنابراین $n-k$ سمبل s_2) وجود داشته باشد آنگاه احتمال s برابر $Pr(\mathbf{s})$ است

$$Pr(\mathbf{s}) = (\frac{2}{3})^k (\frac{1}{3})^{n-k} = \frac{2^k}{3^n}.$$

برای هر $0 \le k \le n$ تعداد این سمبل‌ها برابر $\binom{n}{k}$ است. در کدگذاری شانون — فانو (بخش ۴.۳) به هر یک چنین سمبل s یک کلمه به طول l_k نسبت داده می‌شود

$$l_k = \lceil \log_2 (\frac{1}{Pr(\mathbf{s})}) \rceil = \lceil \log_2 (\frac{3^n}{2^k}) \rceil = \lceil n \log_2 3 - k \rceil = a_n - k,$$

که a_n معرف $\lceil n \log_2 3 \rceil$ است. پس متوسط طول کلمه در S^n برابر است با

$$\begin{aligned} L_n &= \sum_{k=0}^{n} \binom{n}{k} Pr(\mathbf{s}) l_k \\ &= \sum_{k=0}^{n} \binom{n}{k} \frac{2^k}{3^n} (a_n - k) \\ &= \frac{1}{3^n} \left(a_n \sum_{k=0}^{n} \binom{n}{k} 2^k - \sum_{k=0}^{n} k \binom{n}{k} 2^k \right). \end{aligned} \qquad (۹.۳)$$

برای محاسبه دو مجموع ظاهر شده در (۹.۳) می‌توان از قضیه دوجمله‌ای استفاده کرد. داریم

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k, \qquad (۱۰.۳)$$

پس با قرار دادن $x = 2$ نتیجه می‌گیریم

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n.$$

معادلاً، عدد $\sum_k \binom{n}{k} \frac{۲^k}{۳^n}$ برابر مجموع احتمال سمبل‌های s در S^n بوده و بنابراین برابر ۱ می‌باشد. با مشتق گرفتن از (۱۰.۳) و سپس ضرب نمودن در x خواهیم داشت

$$nx(۱+x)^{n-۱} = \sum_{k=۱}^{n} k\binom{n}{k}x^k = \sum_{k=۰}^{n} k\binom{n}{k}x^k;$$

پس مجدداً با قرار دادن $x = ۲$ نتیجه می‌گیریم

$$\sum_{k=۰}^{n} k\binom{n}{k}۲^k = ۲n.۳^{n-۱}.$$

از جایگزینی در (۹.۳) خواهیم داشت

$$L_n = \frac{۱}{۳^n}(a_n ۳^n - ۲n.۳^{n-۱}) = a_n - \frac{۲n}{۳},$$

و از این‌رو

$$\frac{L_n}{n} = \frac{a_n}{n} - \frac{۲}{۳} = \frac{\lceil n\log_۲ ۳\rceil}{n} - \frac{۲}{۳}.$$

حال از $n\log_۲ ۳ + ۱ > \lceil n\log_۲ ۳\rceil \geq ۳ \log_۲ n$ نتیجه می‌شود

$$\log_۲ ۳ \leq \frac{\lceil n\log_۲ ۳\rceil}{n} < \frac{۱}{n} + \log_۲ ۳,$$

و بنابراین

$$\frac{\lceil n\log_۲ ۳\rceil}{n} \longrightarrow \log_۲ ۳.$$

پس اگر $n \longrightarrow \infty$ آنگاه

$$\frac{L_n}{n} \longrightarrow \log_۲ ۳ - \frac{۲}{۳}.$$

این حد برابر $H_۲(S) \approx ۰.۹۱۸$ است، پس برای این منبع خاص قضیه شانون را تصدیق کرده‌ایم.

برای $n = ۱,\ldots,۱۰$ میانگین طول کد $L = L_n/n$ به همراه کارایی $\eta = H/L$ در جدول زیر ارائه شده است. (از تقریب $\log_۲ ۳ \approx ۱.۵۸۵$ برای محاسبه $a_n = \lceil n\log_۲ ۳\rceil$ استفاده شده است.)

n	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
a_n	۲	۴	۵	۷	۸	۱۰	۱۲	۱۳	۱۵	۱۶
L	۱.۳۳۳	۱.۳۳۳	۱	۱.۰۸۳	۰.۹۳۳	۱	۱.۰۴۸	۰.۹۵۸	۱	۰.۹۳۳
η	۰.۶۸۹	۰.۶۸۹	۰.۹۱۸	۰.۸۴۸	۰.۹۸۴	۰.۹۱۸	۰.۸۷۶	۰.۹۵۹	۰.۹۱۸	۰.۹۸۴

این نشان می‌دهد $\eta \longrightarrow ۱$ (یعنی این‌که $L \longrightarrow H$) وقتی $n \longrightarrow \infty$، اگرچه این تقارب نسبتاً کند و نامنظم است. اگر به‌عوض کدهای شانون — فانو از کدهای هافمن استفاده کنیم جدول زیر برای $n \leq ۵$ حاصل می‌شود.

n	۱	۲	۳	۴	۵
L	۱	۰.۹۴۴	۰.۹۳۸	۰.۹۳۸	۰.۹۲۳
η	۰.۹۱۸	۰.۹۷۲	۰.۹۷۹	۰.۹۷۹	۰.۹۹۵

در این حالت تقارب η به ۱ نسبتاً سریع‌تر صورت می‌گیرد، اگرچه برای مقادیری خاص از n، مانند $n = ۵$، کدگذاری شانون — فانو تقریباً کارایی روش هافمن را دارد. برای حالت $n = ۵$ این اتفاق به جهت $۲^۸ = ۲۵۶ \approx ۲۴۳ = ۳^n$ می‌باشد؛ بنابراین معکوس احتمال سمبل‌های $\mathcal{S}^۵$ تنها کمی کوچک‌تر از توان‌های ۲ بوده و از این‌رو گرد کردن لگاریتم آن‌ها توسط تابع سقف تنها تأثیر کمی در بر دارد.

تمرین ۵.۳ آنتروپی دوتایی $H_۲(\mathcal{S})$ را پیدا کنید که $\mathcal{S}$ دارای دو سمبل با احتمال‌های $\frac{۴}{۵}$ و $\frac{۱}{۵}$ است. میانگین طول کلمه L_n را برای یک کد شانون — فانو برای منبع $\mathcal{S}^n$ پیدا کرده و نشان دهید که اگر $n \longrightarrow \infty$ آنگاه $\frac{۱}{n}L_n \longrightarrow H_۲(\mathcal{S})$.

تمرین ۶.۳ فرض کنید $\mathcal{S}$ دارای q سمبل با توزیع یکسان باشد. متوسط طول کد، L_n، را برای یک کد r–تایی شانون — فانو برای $\mathcal{S}^n$ پیدا کرده و تصدیق کنید که اگر $n \longrightarrow \infty$ آنگاه $\frac{۱}{n}L_n \longrightarrow H_r(\mathcal{S})$.

۸.۳ تمرینات تکمیلی

تمرین ۷.۳ فرض کنید $f : (۰, ۱] \longrightarrow \mathbf{R}$ یک تابع اکیداً نزولی بوده و برای هر $a, b \in (۰, ۱]$ رابطه $f(ab) = f(a) + f(b)$ برقرار باشد. نشان دهید که برای عددی چون $r > ۱$ داریم $f(x) = -\log_r x$؛ این توجیه‌ای برای تعریف تابع I در (۱.۳) ارائه می‌دهد. (راهنمایی: تابع $g(x) = f(e^{-x})$ را برای $x \geq ۰$ درنظر بگیرید.)

تمرین ۸.۳ منبع $\mathcal{S}$ معرف مجموع نتایج پرتاب دو تاس نااریب مستقل می‌باشد. توزیع احتمال و آنتروپی دوتایی $\mathcal{S}$ را مشخص کنید. میانگین طول کلمه کدهای دوتایی هافمن و شانون — فانو را برای منبع $\mathcal{S}$ به‌دست آورید.

تمرین ۹.۳ نمودار توابع $-p\log p$، $-\bar{p}\log\bar{p}$ و $H(p) = -p\log p - \bar{p}\log\bar{p}$ را برای $۰ \leq p \leq ۱$ رسم کنید که در آن $\log = \log_۲$. نمودار توابع $p\lceil -\log p\rceil$، $\bar{p}\lceil -\log\bar{p}\rceil$ و $p\lceil -\log p\rceil + \bar{p}\lceil -\log\bar{p}\rceil$ (که متوسط طول کد $L(C)$ از کد دوتایی شانون — فانو برای $\mathcal{S}$ است) را رسم کنید. گراف‌های $H(p)$، $L(C)$، و $H(p) + ۱$ را در یک نمودار رسم کرده و درستی قضیه ۱۶.۳ را برای این منبع ملاحظه کنید.

تمرین ۱۰.۳ نشان دهید اگر $q \geq ۲$ آنگاه یک منبع $\mathcal{S}$ با q سمبل و یک کد r–تایی $\mathcal{C}$ برای $\mathcal{S}$ با خاصیت $L(\mathcal{C}) = H_r(\mathcal{S})$ وجود دارند اگر و فقط اگر $q = ۱ \bmod (r - ۱)$.

تمرین ۱۱.۳ آنتروپی سه‌تایی $H_۳(\mathcal{S})$ را برای منبع دو عضوی $\mathcal{S}$ با احتمال‌های $\frac{۳}{۴}$ و $\frac{۱}{۴}$ مشخص کنید. میانگین طول کلمه L_n را برای یک کد سه‌تایی شانون – فانو برای منبع $\mathcal{S}^n$ پیدا کرده و تصدیق کنید اگر $n \longrightarrow \infty$ آنگاه $\frac{۱}{n}L_n \longrightarrow H_۳(\mathcal{S})$. آیا محاسبات مشابه‌ای برای کدهای شانون – فانو دوتایی برای $\mathcal{S}^n$ برقرار است؟

تمرین ۱۲.۳ نشان دهید اگر $q \geq ۲$، $r \geq ۲$ و $\varepsilon > ۰$، آنگاه یک منبع $\mathcal{S}$ با احتمال‌های مثبت وجود دارد به‌قسمی‌که هر کد لحظه‌ای $\mathcal{C}$ برای $\mathcal{S}$ در رابطه $L(\mathcal{C}) > ۱ + H_r(\mathcal{S}) - \varepsilon$ صدق می‌کند.

تمرین ۱۳.۳ آنتروپی r تایی یک منبع $\mathcal{S}$ که حاوی نامتناهی سمبل با احتمال‌های $p_k (k = ۱, ۲, ۳, \ldots)$ می‌باشد را چگونه تعریف می‌کنید؟ مقدار $H_۲(\mathcal{S})$ را محاسبه کنید که $p_k = ۲^{-k}$، و یک کد دوتایی لحظه‌ای برای این منبع پیدا کنید که میانگین طول کلمه آن برابر $H_۲(\mathcal{S})$ باشد.

تمرین ۱۴.۳ یک منبع $\mathcal{S}$ با خروجی $X_۱, X_۲, \ldots$ از سمبل‌های $s_i \in S$ تشکیل یک منبع مارکف با حافظه ۱ –سمبل می‌دهد؛ این بدین معنی است که احتمال‌های شرطی و ثابت $p_{ij} = Pr(X_{n+۱} = s_j | X_n = s_i)$ مستقل از n داده شده‌اند. اگر فرض کنیم که هر p_{ij} عددی مثبت است آنگاه می‌توان نشان داد که در یک پریود طولانی سمبل‌های s_i احتمال‌های ثابت و مثبتی خواهند داشت که با $p_i > ۰$ نمایش می‌دهیم. تعریف $H(\mathcal{S}) = -\sum_i \sum_j p_i p_{ij} \log p_{ij}$ را که آنتروپی $\mathcal{S}$ می‌باشد توضیح دهید. ثابت کنید $H(\mathcal{S}) \leq H(\mathcal{T})$ که در آن $\mathcal{T}$ یک منبع بدون حافظه با سمبل‌های s_i و احتمال‌های p_i می‌باشد و تعیین کنید که چه موقع تساوی رخ می‌دهد. تفسیر این نتیجه چیست؟ مقادیر $H(\mathcal{S})$، $H(\mathcal{T})$، و (p_i) را مشخص کنید که در آن احتمال‌های p_{ij} با ماتریس زیر مشخص می‌شوند

$$(p_{ij}) = \frac{۱}{۶} \begin{pmatrix} ۳ & ۲ & ۱ \\ ۱ & ۴ & ۱ \\ ۱ & ۲ & ۳ \end{pmatrix}.$$

فصل ۴

کانال‌های اطلاعات

سخنان دو پهلوی اعجوبه‌ای که دروغ می‌گوید به راستی شباهت دارد. (Macbeth)

در این فصل یک منبعی را درنظر می‌گیریم که از طریق یک کانال ناامن (یا پارازیت‌دار) به یک گیرنده پیام می‌فرستد. پارازیت کانال می‌تواند خطاهای مکانیکی و یا انسانی، و یا تداخلات ایجاد شده به‌وسیله منابع دیگر باشد. یک مثال خوب می‌تواند یک آزمایشگاه فضایی باشد که ذخیره توانی محدود داشته و باید پیام‌هایی ارسال کند که علی‌رغم وجود بسیاری از سیگنال‌های قوی‌تر رقابت کننده باید قابل دریافت باشند. به‌دلیل وجود پارازیت سیگنال‌های دریافتی ممکن است همان سیگنال‌های ارسال شده نباشند. هدف ما این است که میزان اطلاعات ارسال شده و از دست رفته را با به‌کار بردن چند نوع مختلف از تابع آنتروپی اندازه‌گیری کرده و سپس آن را با میانگین طول کد به‌کار رفته مرتبط سازیم.

۱.۴ نمادها و تعاریف

ورودی یک کانال اطلاعاتی Γ را یک منبع A با الفبای متناهی A از سمبل‌های a_1، ...، a_r و با احتمال‌های $p_i = Pr(a = a_i)$ درنظر می‌گیریم. طبق معمول

$$0 \leq p_i \leq 1, \quad \sum_{i=1}^{r} p_i = 1.$$

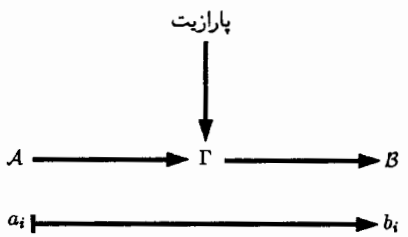
در این‌جا A می‌تواند یک منبع S، با $a_i = s_i$ (سمبل‌های منبع) باشد، و یا معادلاً A می‌تواند معرف منبع S به همراه یک کد C برای S باشد، که در این حالت سمبل‌های a_i می‌توانند

معرف کدسمبل‌های t_i یا کدکلمه‌های w_i باشند. به‌منظور ایجاد فضا برای تمامی این تعبیرات ما نمادها را به A، A و a_i تغییر داده‌ایم.

فرض بر این است که هر وقت یک سمبل $a_i \in A$ به توی کانال Γ فرستاده شود یک سمبلی از Γ ظاهر می‌شود. خروجی Γ به‌عنوان منبع $\mathcal{B}$ با الفبای متناهی B متشکل از سمبل‌های b_1، ...، b_s با احتمال‌های $q_j = Pr(b = b_j)$ درنظر گرفته می‌شود که

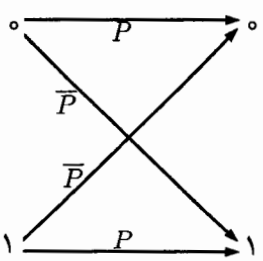
$$0 \leq q_j \leq 1, \quad \sum_{i=1}^{s} q_j = 1.$$

شکل ۱.۴ این موقعیت را توضیح می‌دهد.



شکل ۱.۴

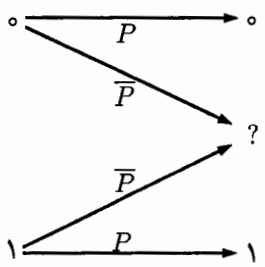
مثال ۱.۴ در کانال دوتایی متقارن[1] (که مختصراً با BSC نشان داده می‌شود) داریم $A = B = \mathbf{Z}_2 = \{0, 1\}$. هر سمبل ورودی $a = 0$ یا $a = 1$ با احتمال P درست و با احتمال $\overline{P} = 1 - P$ غلط (به‌فرم $\bar{a} = 1 - a$) ارسال می‌شود، که P عددی بین ۰ و ۱ است، $0 \leq P \leq 1$. این با شکل ۲.۴ توضیح داده شده است.



شکل ۲.۴

مثال ۲.۴ در کانال پاک‌کننده دوتایی[2] (BEC) داریم $A = \mathbf{Z}_2 = \{0, 1\}$ و $B = \{0, 1, ?\}$. هر سمبل ورودی a با احتمال P درست ارسال شده، و با احتمال $\bar{P}$ پاک (یا غیر قابل تشخیص) می‌شود که با سمبل $b =$? نشان داده می‌شود (شکل ۳.۴ را ببینید).

[1] Binary symmetric channel [2] Binary erasure channel



شکل ۳.۴

در حالت کلی فرض خواهیم کرد که رفتار Γ با احتمال‌های به‌جلو خود کاملاً مشخص شود

$$P_{ij} = Pr(b = b_j | a = a_i) = Pr(b_j | a_i).$$

بنابراین P_{ij} یک احتمال شرطی است که براساس آن به‌فرض این‌که سمبل ورودی a_i باشد احتمال این‌که سمبل خروجی b_j باشد برابر با P_{ij} است. فرض می‌کنیم که P_{ij} مستقل از زمان و نیز مستقل از تمامی سمبل‌های قبلاً ارسال شده و دریافت شده باشد. اگر $a = a_i$، آنگاه b باید دقیقاً یکی از سمبل‌های خروجی b_j باشد، پس برای هر $1 \leq i \leq r$:

$$\sum_{j=1}^{s} P_{ij} = 1.$$

این اعداد P_{ij} ماتریس کانال را تشکیل می‌دهند

$$M = (P_{ij}) = \begin{pmatrix} P_{11} & \cdots & P_{1s} \\ \vdots & & \vdots \\ P_{r1} & \cdots & P_{rs} \end{pmatrix},$$

که دارای r سطر و s ستون می‌باشد. عنصر سطر i ام و ستون j ام برابر P_{ij} است. به‌عنوان نمونه اگر Γ کانال BSC یا BEC باشد آنگاه داریم

$$M = \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix} \quad \text{یا} \quad \begin{pmatrix} P & 0 & \overline{P} \\ 0 & P & \overline{P} \end{pmatrix}.$$

شکل دقیق ماتریس کانال بستگی به ترتیب سمبل‌های ورودی a_i و خروجی b_j دارد؛ یک تغییر ترتیب منجر به یک جایگشت سطری یا ستونی می‌شود. بنابراین ماتریسی که برای BEC ارائه شد از ترتیب ۰، ۱، ؟ در سمبل‌های خروجی استفاده می‌کند، حال آن‌که اگر ترتیب ۰، ؟، ۱ را درنظر بگیریم ماتریس زیر حاصل می‌شود

$$M = \begin{pmatrix} P & \overline{P} & 0 \\ 0 & \overline{P} & P \end{pmatrix}.$$

به چند روش می‌توان دو کانال Γ و Γ' را ترکیب کرده و کانال سومی را به‌دست آورد. اگر Γ و Γ' دارای الفبای ورودی متمایز A و A' و الفبای خروجی متمایز B و B' باشند آنگاه مجموع $\Gamma + \Gamma'$ دارای الفبای ورودی و خروجی $A \cup A'$ و $B \cup B'$ می‌باشد: هر سمبل ورودی از طریق Γ یا Γ' ارسال می‌شود زیرا آن سمبل در A یا A' قرار دارد. از این‌رو ماتریس کانال ماتریس قالبی

$$\begin{pmatrix} M & O \\ O & M' \end{pmatrix}$$

است که M و M' به‌ترتیب ماتریس کانال برای Γ و Γ' می‌باشند. این‌را می‌توان به‌سادگی به مجموع هر تعداد متناهی از کانال تعمیم داد.

در حالت $\Gamma \times \Gamma'$ نیازی به فرض متمایز بودن A و A' یا B و B' نیست. الفبای ورودی و خروجی به‌ترتیب برابر $A \times A'$ و $B \times B'$ می‌باشند، و فرستنده یک زوج $(a, a') \in A \times A'$ را با ارسال هم‌زمان a از طریق Γ و a' از طریق Γ' ارسال می‌کند. در این وضعیت یک زوج $(b, b') \in B \times B'$ دریافت می‌شود. بنابراین احتمال‌های به‌جلو عبارت هستند از

$$Pr((b, b')|(a, a')) = Pr(b|a).Pr(b'|a').$$

پس در این‌جا ماتریس کانال ضرب کرانکر[1] ماتریس‌های M و M' است که با $M \otimes M'$ نشان داده می‌شود. اگر $M = (P_{ij})$ و $M' = (P'_{kl})$ به‌ترتیب دو ماتریس $r \times s$ و $r' \times s'$ باشند آنگاه $M \otimes M'$ یک ماتریس $rr' \times ss'$ با درایه‌های $P_{ij}P'_{kl}$ است. (ترتیب اعضای این ماتریس بستگی به ترتیب در $A \times A'$ و $B \times B'$ دارد.) در این حالت هم می‌توان تعریف را به ضرب یک تعداد متناهی از کانال تعمیم داد، بالاخص بسط n ام یک کانال Γ که با Γ^n نمایش داده می‌شود ضرب n کپی از Γ بوده و $\Gamma^n = \Gamma \times \cdots \times \Gamma$.

مثال ۳.۴ اگر Γ و Γ' کانال‌های متقارن دوتایی با ماتریس کانال M و M' باشند

$$M = \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix}, \quad M' = \begin{pmatrix} P' & \overline{P'} \\ \overline{P'} & P' \end{pmatrix},$$

آنگاه $\Gamma + \Gamma'$ و $\Gamma \times \Gamma'$ به‌ترتیب دارای ماتریس‌های M_1 و M_2 می‌باشند:

$$M_1 = \begin{pmatrix} P & \overline{P} & 0 & 0 \\ \overline{P} & P & 0 & 0 \\ 0 & 0 & P' & \overline{P'} \\ 0 & 0 & \overline{P'} & P' \end{pmatrix}, \quad M_2 = \begin{pmatrix} PP' & \overline{P}P' & P\overline{P'} & \overline{P}\,\overline{P'} \\ \overline{P}P' & PP' & \overline{P}\,\overline{P'} & P\overline{P'} \\ P\overline{P'} & \overline{P}\,\overline{P'} & PP' & \overline{P}P' \\ \overline{P}\,\overline{P'} & P\overline{P'} & \overline{P}P' & PP' \end{pmatrix}.$$

[1] Kronecker product

در ارتباط با $\Gamma \times \Gamma'$، برای $A \times A' = B \times B' = \mathbf{Z}_2^2$ ترتیب $(0,0)$، $(1,0)$، $(0,1)$، $(1,1)$ را درنظر گرفته‌ایم.

تمرین ۱.۴ خروجی یک کانال Γ به‌عنوان ورودی کانال Γ' منظور می‌شود. یک ماتریس کانال برای کانال مرکب $\Gamma \circ \Gamma'$ برحسب ماتریس کانال‌های Γ و Γ' ارائه دهید. این نتیجه را به ترکیب هر تعداد متناهی از کانال‌های سری تعمیم دهید (این یک آبشار از کانال‌ها نامیده می‌شود).

با برگشت به حالت کانال منفرد Γ، اگر معادلات $\sum_i p_i = 1$ و $\sum_i P_{i,j} = 1$ را در هم ضرب کنیم نتیجه می‌شود

$$\sum_{i=1}^{r}\sum_{j=1}^{s} p_i P_{ij} = 1. \qquad (1.4)$$

احتمال این‌که a_i فرستاده شده و b_j دریافت شود برابر است با $p_i P_{ij}$. اگر b_j دریافت شود آنگاه دقیقاً یکی از سمبل‌های a_i باید ارسال شده باشد، پس برای هر $1 \le j \le s$ داریم

$$\sum_{i=1}^{r} p_i P_{ij} = q_j. \qquad (2.4)$$

اگر (p_i) را به‌عنوان یک بردار $\mathbf{p} \in \mathbf{R}^r$ و (q_j) را به‌عنوان یک بردار $\mathbf{q} \in \mathbf{R}^s$ درنظر بگیریم آنگاه (۲.۴) را می‌توان به‌فرم زیر نوشت

$$\mathbf{p}M = \mathbf{q}. \qquad (2'.4)$$

اگر در (۲.۴) روی j عمل جمع انجام داده و سپس ترتیب جمع را عوض کرده و از $\sum_j q_j = 1$ استفاده کنیم (۱.۴) حاصل می‌شود.

علاوه بر احتمال‌های به‌جلو P_{ij}، تعریف احتمال‌های به‌عقب نیز مفید است که چنین تعریف می‌شود

$$Q_{ij} = Pr(a = a_i | b = b_j) = Pr(a_i | b_j).$$

احتمال‌های توام نیز تعریف می‌شوند:

$$R_{ij} = Pr(a = a_i, b = b_j) = Pr(a_i, b_j).$$

احتمال‌های به‌جلو P_{ij} را می‌توان معرف دیدگاه فرستنده درنظر گرفت که سمبل ارسالی a_i را می‌داند و تلاش می‌کند سمبل خروجی را حدس بزند. مشابهاً احتمال‌های به‌عقب

دیدگاه دریافت کننده را منعکس می‌سازد که خروجی را دانسته و تلاش در برآوردی از سمبل ورودی a_i دارد، درحالی‌که احتمال‌های توام R_{ij} یک ناظر خارجی را معرفی می‌کند که تلاش می‌کند ورودی a_i و خروجی b_j را حدس بزند.

برای هر i و j داریم

$$p_iP_{ij} = Pr(a_i)Pr(b_j|a_i) = Pr(a_i,b_j) = Pr(b_j)Pr(a_i|b_j) = q_jQ_{ij},$$

و از این‌جا فرمول بیز[1] مشروط به $q_j \neq 0$ به‌دست می‌آید

$$Q_{ij} = \frac{p_i}{q_j}P_{ij}. \qquad (۳.۴)$$

از ترکیب (۳.۴) و (۲.۴) نتیجه می‌شود

$$Q_{ij} = \frac{p_iP_{ij}}{\sum_{k=1}^{r} p_kP_{kj}}. \qquad (۴.۴)$$

در بخش بعد مثال‌های خاصی از این معادلات را ملاحظه خواهیم کرد.

۲.۴ کانال دوتایی متقارن

یکی از ساده‌ترین کانال‌های اطلاعات که اکثراً به‌کار می‌رود کانال دوتایی متقارن (BSC) است که در مثال ۱.۴ معرفی شد. به جهت اهمیت آن در این قسمت به مطالعه دقیق‌تر آن می‌پردازیم. به‌خاطر بیاورید که این کانال چنین تعریف می‌شود:

۱) $A = B = \mathbf{Z}_2 = \{0,1\}$،

۲) به‌ازای عددی چون $0 \leq P \leq 1$، ماتریس کانال به‌فرم زیر است:

$$M = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix} = \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix}.$$

برای سادگی در نمادگذاری، در این‌جا از اندیس‌های $i,j = 0,1$ به‌جای ۱ و ۲ استفاده می‌کنیم، از این‌رو بر اساس نماد بخش ۱.۴ داریم $a_i = i$ و $b_j = j$.

شرط (۱) به دوتایی بودن Γ و شرط (۲) به متقارن بودن آن اشاره می‌کند، بدین معنی که هر سمبل ورودی a با احتمال P درست و با احتمال $\overline{P}$ غلط ارسال می‌شود. احتمال‌های ورودی عبارت هستند از

[1] Bayes's formula

$$p_۰ = Pr(a = ۰) = p,$$
$$p_۱ = Pr(a = ۱) = \overline{p},$$

که p عددی در $[۰, ۱]$ است. در این‌صورت روابط (۲.۴) به شرح زیر خواهد بود

$$q_۰ = Pr(b = ۰) = pP + \overline{p}\,\overline{P},$$
$$q_۱ = Pr(b = ۱) = p\overline{P} + \overline{p}P;$$

با نوشتن $q_۰ = q$ و $q_۱ = \overline{q}$، همچون تساوی (۲′.۴) رابطه زیر حاصل می‌شود

$$(q, \overline{q}) = (p, \overline{p}) \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix}.$$

از جایگزین کردن این مقادیر در فرمول بیز (۳.۴) نتیجه می‌شود:

$$Q_{۰۰} = \frac{pP}{pP + \overline{p}\,\overline{P}}, \qquad Q_{۱۰} = \frac{\overline{p}\,\overline{P}}{pP + \overline{p}\,\overline{P}},$$
$$Q_{۰۱} = \frac{p\,\overline{P}}{p\overline{P} + \overline{p}P}, \qquad Q_{۱۱} = \frac{\overline{p}P}{p\overline{P} + \overline{p}P}.$$

مثال ۴.۴ فرض کنید ورودی $\mathcal{A}$ با $p = \frac{۱}{۲}$ تعریف شده باشد. پس $\overline{p} = \frac{۱}{۲}$، و بنابراین سمبل‌های ورودی $a = ۰, ۱$ توزیع یکنواخت دارند. داریم $q = \frac{۱}{۲}P + \frac{۱}{۲}\overline{P} = \frac{۱}{۲}(P + \overline{P}) = \frac{۱}{۲}$ و مشابهاً $\overline{q} = \frac{۱}{۲}$، پس سمبل‌های خروجی $b = ۰, ۱$ نیز توزیع یکنواخت دارند. احتمال‌های به‌عقب برابر هستند با:

$$Q_{۰۰} = Q_{۱۱} = \frac{\frac{۱}{۲}P}{\frac{۱}{۲}} = P, \qquad Q_{۰۱} = Q_{۱۰} = \frac{\frac{۱}{۲}\overline{P}}{\frac{۱}{۲}} = \overline{P}.$$

مثال ۵.۴ فرض کنید $P = ۰.۸$ (پس Γ نسبتاً قابل اعتماد است، از هر ۱۰ سمبل ارسال شده هشت سمبل درست دریافت می‌شود)، و $p = ۰.۹$ (از این‌رو سمبل ورودی تقریباً همیشه ۰ است). در این‌صورت ملاحظه می‌کنیم که

$$q_۰ = q = pP + \overline{p}\,\overline{P} = ۰.۷۴, \qquad q_۱ = \overline{q} = p\overline{P} + \overline{p}P = ۰.۲۶.$$

بنابراین سمبل خروجی b معمولاً ۰ است، ولی تمایل به ۰ به‌اندازه تمایل ورودی به ۰ نیست. این کاهش در تمایل به ۰ به‌دلیل پارازیت (خطاهای کانال) است. براین اساس تعداد دفعاتی که یک ۰ ارسالی تبدیل به ۱ می‌شود بیشتر از دفعاتی است که یک ۱ ارسال شده

به‌عنوان ۰ دریافت می‌شود؛ این صرفاً به این دلیل است که سمبل‌های بیشتری از ۰ ارسال می‌شود. احتمال‌های به‌عقب عبارت هستند از

$$Q_{00} = \frac{p_0 P_{00}}{q_0} = \frac{0.9 \times 0.8}{0.74} \approx 0.973, \quad Q_{10} = \frac{p_1 P_{10}}{q_0} = \frac{0.1 \times 0.2}{0.74} \approx 0.027,$$
$$Q_{01} = \frac{p_0 P_{01}}{q_1} = \frac{0.9 \times 0.2}{0.26} \approx 0.692, \quad Q_{11} = \frac{p_1 P_{11}}{q_1} = \frac{0.1 \times 0.8}{0.26} \approx 0.308.$$

پس اگر $b = 0$ آنگاه تقریباً همیشه $a = 0$ بوده و اگر $b = 1$ آنگاه معمولاً $a = 0$ است. بنابراین صرف‌نظر از سمبل دریافتی، سمبل ورودی با احتمال بالا ۰ بوده است. کانال BSC رفتاری مشابه دارد هرگاه $Q_{00} > Q_{10}$ و $Q_{01} > Q_{11}$، یعنی این‌که $pP > \overline{p}\overline{P} = (1-p)\overline{P}$ و $p\overline{P} > \overline{p}P = (1-p)P$؛ این دو نامساوی را می‌توان به‌فرم $p = p(P+\overline{P}) > \overline{P}$ و $p = p(\overline{P}+P) > P$، و یا معادلاً به‌صورت $p > \max(P, \overline{P})$ نوشت. مشابهاً، اگر $\overline{p} > \max(P, \overline{P})$ آنگاه سمبل ارسالی با احتمال بالا ۱ خواهد بود.

تمرین ۲.۴ فرض کنید Γ کانال BSC باشد. شرایط لازم و کافی، روی p و P، ارائه دهید که Γ در روابط زیر صدق کند

(i) $Q_{00} < Q_{10}$ و $Q_{01} < Q_{11}$؛

(ii) $Q_{00} > Q_{10}$ و $Q_{01} < Q_{11}$؛

(iii) $Q_{00} < Q_{10}$ و $Q_{01} > Q_{11}$.

از دیدگاه دریافت کننده معنی این شرایط چیست؟

۳.۴ دستگاه آنتروپی

در کانال Γ ورودی $\mathcal{A}$ و خروجی $\mathcal{B}$ منابعی هستند که آنتروپی‌های مختص خود را دارند؛

$$H(\mathcal{A}) = \sum_i p_i \log \frac{1}{p_i}, \qquad H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}.$$

این‌ها معرف میانگین اطلاعات وارد شده و خارج شده از کانال Γ به‌ازای هر سمبل، یا معادلاً ابهام ما نسبت به ورودی و خروجی، هستند.

با فرض این‌که $b = b_j$ دریافت شده باشد یک آنتروپی شرطی وجود دارد

$$H(\mathcal{A}|b_j) = \sum_i Pr(a_i|b_j) \log \frac{1}{Pr(a_i|b_i)} = \sum_i Q_{ij} \log \frac{1}{Q_{ij}};$$

این نمایش‌گر ابهام دریافت کننده نسبت به $\mathcal{A}$ است وقتی b_j دریافت شده است، به عبارت دیگر این عدد میزان اطلاعات اضافی را که از دانستن $\mathcal{A}$ حاصل می‌شود تعیین می‌کند. با میانگین گرفتن روی b_j و به‌کار بردن $R_{ij} = q_j Q_{ij}$، ایهام $\mathcal{A}$ نسبت به $\mathcal{B}$ را به‌دست می آوریم

$$H(\mathcal{A}|\mathcal{B}) = \sum_j q_j H(\mathcal{A}|b_j) = \sum_j q_j \left(\sum_i Q_{ij} \log \frac{1}{Q_{ij}} \right) = \sum_i \sum_j R_{ij} \log \frac{1}{Q_{ij}}.$$

این معرف میانگین ابهام دریافت کننده راجع به $\mathcal{A}$ است وقتی $\mathcal{B}$ دریافت شده است؛ معادلاً این میزان اطلاعات اضافی حاصل از دانستن $\mathcal{A}$ را تعیین می‌کند. مشابهاً، اگر a_i ارسال شود آنگاه ابهام راجع به $\mathcal{B}$ برابر آنتروپی شرطی $H(\mathcal{B}|a_i)$ می‌باشد

$$H(\mathcal{B}|a_i) = \sum_j Pr(b_j|a_i) \log \frac{1}{Pr(b_j|a_i)} = \sum_j P_{ij} \log \frac{1}{P_{ij}}.$$

با میانگین گرفتن روی a_i و به‌کار بردن $R_{ij} = p_i P_{ij}$ مقدار $H(\mathcal{B}|\mathcal{A})$ مشخص می‌شود

$$H(\mathcal{B}|\mathcal{A}) = \sum_i p_i H(\mathcal{B}|a_i) = \sum_i p_i (\sum_j P_{ij} \log \frac{1}{P_{ij}}) = \sum_i \sum_j R_{ij} \log \frac{1}{P_{ij}};$$

این ایهام $\mathcal{B}$ نسبت به $\mathcal{A}$ بوده و معرف میانگین ابهام فرستنده راجع به خروجی $\mathcal{B}$ است وقتی‌که $\mathcal{A}$ معلوم است. معادلاً این عدد برابر میزان اطلاعات اضافی حاصل از دانستن $\mathcal{B}$، مشروط به معلوم بودن $\mathcal{A}$، است.

یک ناظری که تلاش می‌کند هر دو ورودی و خروجی کانال Γ را حدس بزند دارای میانگین ابهام مشخص‌شده با آنتروپی توام می‌باشد

$$H(\mathcal{A}, \mathcal{B}) = \sum_i \sum_j Pr(a_i, b_j) \log \frac{1}{Pr(a_i, b_j)} = \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}}.$$

اگر $\mathcal{A}$ و $\mathcal{B}$ از نظر آماری مستقل باشند، یعنی این‌که برای هر i و j تساوی $R_{ij} = p_i q_j$ برقرار باشد (چیزی که در زندگی واقعی خیلی محتمل نیست!)، آنگاه داریم

$$\begin{aligned} H(\mathcal{A}, \mathcal{B}) &= \sum_i \sum_j p_i q_j (\log \frac{1}{p_i} + \log \frac{1}{q_j}) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} \qquad (\text{چون } \textstyle\sum_i p_i = \sum_j q_j = 1) \\ &= H(\mathcal{A}) + H(\mathcal{B}). \qquad (۵.۴) \end{aligned}$$

بنابراین در این حالت اطلاعات حمل‌شده توسط $\mathcal{A}$ و $\mathcal{B}$ برابر است با مجموع اطلاعات حمل‌شده توسط هر یک از آن‌ها (در حالات دیگر، خواهیم دید که این مقدار کمتر از مجموع است). اگر آنتروپی را به‌عنوان مشخص کننده مقدار اطلاعات (یا ابهام) درنظر بگیریم، آنگاه (۵.۴) شبیه به نتیجه $|A \cup B| = |A| + |B|$ برای دو مجموعه متناهی و متمایز A و B می‌باشد.

در حالت کلی انتظار می‌رود که $\mathcal{A}$ و $\mathcal{B}$ مرتبط باشند، از این‌رو در چنین حالاتی از $R_{ij} = p_i P_{ij}$ استفاده کرده و نتیجه می‌گیریم

$$H(\mathcal{A},\mathcal{B}) = \sum_i \sum_j R_{ij} \log \frac{1}{p_i} + \sum_i \sum_j R_{ij} \log \frac{1}{P_{ij}}.$$

حال با درنظر گرفتن $\sum_j R_{ij} = p_i$ تساوی زیر به‌دست می‌آید

$$\begin{aligned} H(\mathcal{A},\mathcal{B}) &= \sum_i p_i \log \frac{1}{p_i} + \sum_i \sum_j R_{ij} \log \frac{1}{P_{ij}} \\ &= H(\mathcal{A}) + H(\mathcal{B}|\mathcal{A}). \end{aligned} \tag{۶.۴}$$

این تاییدی بر تفسیر $H(\mathcal{B}|\mathcal{A})$ به‌عنوان اطلاعات اضافی حمل‌شده به‌وسیله $\mathcal{B}$ وقتی‌که $\mathcal{A}$ از قبل معلوم است می‌باشد. این قاعده مشابه با فرمول $|A \cup B| = |A| + |B \backslash A|$ برای دو مجموعه متناهی A و B است. از تعویض نقش $\mathcal{A}$ و $\mathcal{B}$ نتیجه می‌شود

$$H(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A}|\mathcal{B}), \tag{۷.۴}$$

که تفسیر مشابه‌ای برای $H(\mathcal{A}|\mathcal{B})$ داشته و این تساوی متناظر با $|A \cup B| = |B| + |A \backslash B|$ است.

اعداد $H(\mathcal{A})$، $H(\mathcal{B})$، $H(\mathcal{A}|\mathcal{B})$، $H(\mathcal{B}|\mathcal{A})$ و $H(\mathcal{A},\mathcal{B})$ را دستگاه آنتروپی می‌نامیم؛ این اعداد بستگی به Γ و $\mathcal{A}$ دارند.

تمرین ۳.۴ معادله (۷.۴) را ثابت کنید: $H(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A}|\mathcal{B})$. این رابطه چه تفسیری را برای $H(\mathcal{A}|\mathcal{B})$ ایجاب می‌کند؟

تمرین ۴.۴ نشان دهید که دستگاه آنتروپی کانال ضربی $\Gamma \times \Gamma'$ از جمع اعداد متناظر برای Γ و Γ' به‌دست می‌آید، درحالی‌که دستگاه آنتروپی Γ^n با n برابر نمودن اعداد مشابه برای Γ مشخص می‌شود. (راهنمایی: بخش ۵.۳ را ببینید.)

۴.۴ دستگاه آنتروپی برای کانال دوتایی متقارن

فرض کنید با نمادگذاری بخش ۲.۴، Γ یک کانال BSC باشد. آنتروپی‌های ورودی و خروجی عبارت هستند از

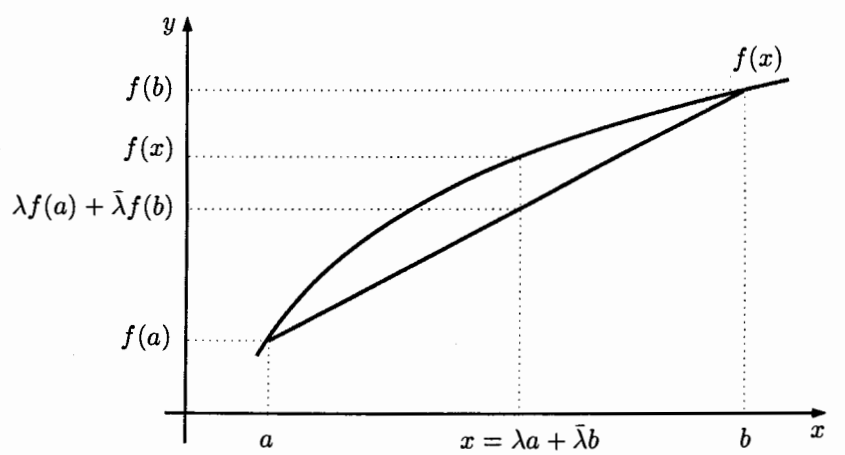
$$\begin{aligned} H(\mathcal{A}) &= -p\log p - \bar{p}\log\bar{p} = H(p), \\ H(\mathcal{B}) &= -q\log q - \bar{q}\log\bar{q} = H(q), \end{aligned}$$

که $q = pP + \bar{p}\bar{P}$. برای مقایسه اینها از تحدب استفاده می‌کنیم.

یک تابع $f : [0,1] \longrightarrow \mathbf{R}$ اکیداً محدب است اگر برای هر $a,b,\lambda \in [0,1]$ و $x = \lambda a + \bar{\lambda} b$ داشته باشیم

$$f(x) \geq \lambda f(a) + \bar{\lambda} f(b),$$

و تساوی برقرار بوده اگر و فقط اگر x برابر a یا b، یعنی $a = b$ یا $\lambda \in \{0,1\}$، باشد. چون $\lambda - 1 = \bar{\lambda}$، x با تغییر λ بین ۰ و ۱ از b تا a در تغییر است. گراف $\lambda f(a) + \bar{\lambda} f(b)$ خط راست واصل دو نقطه $(a, f(a))$ و $(b, f(b))$ است، از این‌رو تحدب بدین معنی است که برای هر نقطه بین a و b مانند x، نمودار $(x, f(x))$ بالای خط واصل یادشده قرار دارد (شکل ۴.۴ را ببینید[1]).



شکل ۴.۴

[1] در بعضی از زمینه‌های ریاضی مانند آنالیز و تحقیق در عملیات نامساوی اصلی در این تعریف معکوس بوده و از این‌رو گراف زیر خط قرار دارد.

گراف تابع $H(p)$، شکل ۳.۳، نشان می‌دهد که این تابع اکیداً محدب است. برای به‌دست آوردن چند نامساوی مهم در رابطه با آنتروپی نیاز به اثبات محدب بودن این تابع داریم. ابتدا نیاز به یک نتیجه کلی از حساب دیفرانسیل می‌باشد.

لم ۶.۴ اگر تابع $f : [0,1] \longrightarrow \mathbf{R}$ پیوسته بوده و به‌ازای هر $x \in (0,1)$ نامساوی $f''(x) < 0$ برقرار باشد آنگاه f اکیداً محدب است.

تمرین ۵.۴ با استفاده از قضیه مقدار میانگین لم ۶.۴ را ثابت کنید. (راهنمایی: این قضیه می‌گوید اگر تابع g روی $[0,1]$ پیوسته و روی $(0,1)$ مشتق‌پذیر باشد، و اگر $0 \leq a < b \leq 1$، آنگاه به‌ازای عددی چون c بین a و b رابطه $g(b) - g(a) = (b-a)g'(c)$ برقرار است؛ به‌عنوان نمونه مرجع [Fi 83] یا [La 83] را ببینید. فرض کنید که لم غلط بوده و با به‌کار بردن قضیه مقدار میانگین دوبار روی f و سپس یک‌بار روی f' یک تناقض به‌دست آورید.)

نتیجه ۷.۴ تابع آنتروپی روی $[0,1]$ اکیداً محدب است.

اثبات برای $0 < p < 1$ رابطه $H(p) = -p\log p - (1-p)\log(1-p)$ برقرار است، پس $H(p)$ پیوسته بوده و روی $(0,1)$ دوبار مشتق‌پذیر است؛ پیوستگی H در 0 و 1 از قرارداد $H(0) = H(1) = 1$ نتیجه می‌شود. بدون از دست دادن کلیت موضوع، می‌توان فرض کرد که لگاریتم‌ها طبیعی هستند، از این‌رو $H'(p) = -\ln p + \ln(1-p)$ و در نتیجه برای هر $p \in (0,1)$ داریم

$$H''(p) = -\frac{1}{p} - \frac{1}{1-p} < 0.$$

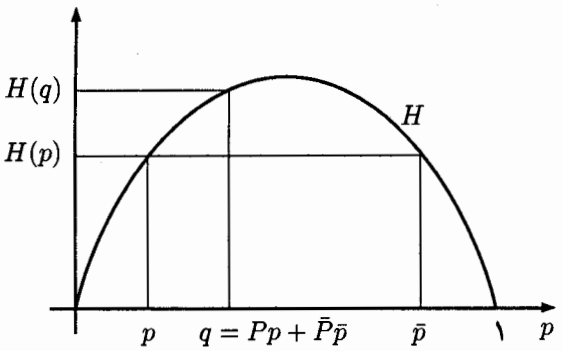
اکنون نتیجه از لم ۶.۴ حاصل می‌شود. ■

تمرین ۱۴.۴ را برای تعمیمی از این نتیجه به منابع با تعداد دلخواه سمبل ملاحظه کنید.

حال به BSC باز می‌گردیم. اگر در تعریف تحدب اکید قرار دهیم $a = p$، $b = \bar{p}$ و $\lambda = P$، می‌بینیم که $x = pP + \bar{p}\bar{P} = q$، و بنابراین $H(q) \geq H(p)$ و تساوی برقرار است اگر و فقط اگر $p = \bar{p}$ (یعنی $p = \frac{1}{2}$) یا $q \in \{p, \bar{p}\}$ (یعنی این‌که $P \in \{0,1\}$). چون $H(\mathcal{A}) = H(p)$ و $H(\mathcal{B}) = H(q)$ پس BSC در شرط

$$H(\mathcal{B}) \geq H(\mathcal{A}) \qquad (۸.۴)$$

صدق کرده و تساوی برقرار است اگر و فقط اگر سمبل‌های ورودی توزیع یکنواخت داشته ($p = \frac{1}{2}$) و یا کانال کاملاً غیرقابل اعتماد ($P = 0$) یا قابل اعتماد ($P = 1$) باشد. نامساوی (۸.۴) که در شکل ۵.۴ تشریح شده است نشان می‌دهد که معمولاً انتقال از طریق BSC باعث افزایش ابهام می‌شود؛ ولی کانال‌هایی هستند که این وضعیت در مورد آنها درست نیست (تمرین ۶.۴).



شکل ۵.۴

تمرین ٦.۴ مثالی از یک کانال اطلاعات Γ با ورودی $\mathcal{A}$ ارائه دهید که برای آن رابطه $H(\mathcal{B}) < H(\mathcal{A})$ برقرار باشد. ($\mathcal{B}$ خروجی کانال است.)

برای کانال BSC داریم

$$\begin{aligned} H(\mathcal{B}|\mathcal{A}) &= \sum_i \sum_j p_i P_{ij} \log \frac{١}{P_{ij}} \\ &= -pP\log P - p\bar{P}\log\bar{P} - \bar{p}\bar{P}\log\bar{P} - \bar{p}P\log P \\ &= -(p+\bar{p})P\log P - (p+\bar{p})\bar{P}\log\bar{P} \\ &= -P\log P - \bar{P}\log\bar{P} \\ &= H(P). \end{aligned}$$

بنابراین ابهام فرستنده درباره خروجی برابر با آنتروپی ارسال درست سمبل‌ها است. این کمترین مقدار خود را به‌ازای $P \in \{٠, ١\}$ اختیار می‌کند (Γ کاملاً غیرقابل اطمینان یا کاملاً قابل اعتماد)، درحالی‌که ماکزیمم آن به‌ازای $P = \frac{١}{٢}$ رخ می‌دهد (یعنی Γ غیر قابل استفاده است).

از معادله (٦.۴) نتیجه می‌شود

$$H(\mathcal{A},\mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}|\mathcal{A}) = H(p) + H(P),$$

درحالی‌که (۷.۴) ایجاب می‌کند

$$H(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A}|\mathcal{B}) = H(q) + H(\mathcal{A}|\mathcal{B}).$$

بنابراین ابهام کانال BSC از رابطه زیر حاصل می‌شود

$$H(\mathcal{A}|\mathcal{B}) = H(p) + H(P) - H(q).$$

چون $0 \leq p \leq 1$، پس q که در رابطه $q = pP + \bar{p}\bar{P}$ صدق می‌کند بین P و $\bar{P}$ قرار داشته و بنابراین بنابه نتیجه ۷.۴ داریم $H(q) \geq H(P)$. بی‌درنگ نتیجه می‌شود که BSC در نامساوی زیر صدق می‌کند

$$H(\mathcal{B}|\mathcal{A}) \leq H(\mathcal{B}); \qquad (۹.۴)$$

در این‌جا تساوی برقرار است اگر و فقط اگر $P = \frac{1}{2}$ یا $p \in \{0, 1\}$. بنابراین ابهام راجع به $\mathcal{B}$ معمولاً با دانستن $\mathcal{A}$ کاهش می‌یابد. مشابهاً، $H(\mathcal{A}|\mathcal{B}) = H(p) + H(P) - H(q) \leq H(p)$ (زیرا $H(q) \geq H(P)$)، پس

$$H(\mathcal{A}|\mathcal{B}) \leq H(\mathcal{A}), \qquad (۱۰.۴)$$

و تساوی برقرار است اگر و فقط اگر $P = \frac{1}{2}$ یا $p \in \{0, 1\}$. این بدین معنی است که ابهام راجع به $\mathcal{A}$ معمولاً با دانستن $\mathcal{B}$ کاهش می‌یابد. بعداً خواهیم دید که نامساوی‌های (۹.۴) و (۱۰.۴) برای تمامی کانال‌ها برقرار هستند.

تمرین ۷.۴ دستگاه آنتروپی را برای کانال دوتایی پاک کننده (BEC) که در بخش ۱.۴ معرفی شد محاسبه کنید. توزیع احتمال برای سمبل‌های ورودی ۰ و ۱ برابر p و $\bar{p}$ فرض می‌شوند. نشان دهید که این کانال در روابط (۹.۴) و (۱۰.۴) صدق می‌کند.

۵.۴ تعمیم قضیه اول شانون به کانال‌های اطلاعات

قضیه اول شانون برای منابع (قضیه ۲۳.۳) می‌گوید بزرگ‌ترین کران پایین برای میانگین طول کلمه در کدهای یکتا دکدپذیر مربوط به منبع $\mathcal{A}$ برابر آنتروپی $H(\mathcal{A})$ است. مشابهاً، قضیه اول شانون برای کانال (که در این بخش ثابت خواهد شد) اظهار می‌دارد که، با فرض دانستن خروجی $\mathcal{B}$، بزرگ‌ترین کران پایین برای متوسط طول کلمه در کدهای یکتا دکدپذیر مربوط به ورودی کانال یعنی $\mathcal{A}$ برابر $H(\mathcal{A}|\mathcal{B})$ است. در هر یک از این دو حالت، می‌نیمم میانگین طول کلمه یک اندازه دقیق برای اطلاعات است. در حالت مربوط به یک کانال Γ، وضعیت را از دید دریافت کننده بررسی می‌کنیم که از $\mathcal{B}$ مطلع بوده ولی نسبت به $\mathcal{A}$ غیرمطمئن است؛ اطلاعات اضافی لازم برای مطمئن شدن نسبت به $\mathcal{A}$ برابر $H(\mathcal{A}|\mathcal{B})$ است و این برابر است با می‌نیمم میانگین طول کلمه که برای حمل اطلاعات اضافی (به‌وسیله دیگری جدای از Γ) لازم است. در واقع، دریافت کننده می‌گوید که '$\mathcal{B}$ را می‌شناسم، ولی نسبت به $\mathcal{A}$ مطمئن نیستم؛ اطلاعات بیشتری به من بدهید تا $\mathcal{A}$ را تشخیص دهم' و ما در صدد اندازه‌گیری این اطلاعات اضافی مورد نیاز هستیم.

فرض کنید $b = b_j$ دریافت شده است؛ با این فرض چگونه می‌توان این اطلاعات اضافی راجع به سمبل ورودی $a = a_i$ را به بهترین فرم کدگذاری کرد؟ همچون فصل ۳، از کدگذاری شانون — فانو برای بسط $\mathcal{A}^n$ از $\mathcal{A}$ استفاده می‌کنیم؛ تنها تفاوت در این است که چون b_j را می‌دانیم از احتمال‌های شرطی $Q_{ij} = Pr(a_i|b_j)$ به‌عوض احتمال‌های غیرشرطی $p_i = Pr(a_i)$ برای $\mathcal{A}$ استفاده می‌کنیم.

برای سادگی ابتدا $n = ۱$ را انتخاب می‌کنیم، پس (با دانستن این‌که $b = b_j$) همچون بخش ۴.۳ یک کد شانون — فانو برای $\mathcal{A}$ می‌سازیم. این یک کد لحظه‌ای r–تایی C_j با متوسط طول کلمه $L_{(j)}$ برای $\mathcal{A}$ است که بنابر قضیه ۱۶.۳ در رابطه زیر صدق می‌کند

$$H(\mathcal{A}|b_j) \le L_{(j)} \le ۱ + H(\mathcal{A}|b_j) \qquad (۱۱.۴)$$

که در آن

$$H(\mathcal{A}|b_j) = \sum_i Q_{ij} \log \frac{۱}{Q_{ij}}.$$

حال یک روش کدگذاری $\mathcal{C}$ برای $\mathcal{A}$ با استفاده از C_j ارائه می‌دهیم به‌قسمی‌که بتوان $\mathcal{C}$ را به‌عنوان میانگینی از کدهای C_j درنظر گرفت. با میانگین گرفتن از جملات (۱۱.۴) روی $b_j \in B$ (با احتمال‌های q_j) می‌بینیم که $\mathcal{C}$ دارای متوسط طول کلمه $L = \sum_j q_j L_{(j)}$ با خاصیت زیر می‌باشد

$$H(\mathcal{A}|\mathcal{B}) \le L \le ۱ + H(\mathcal{A}|\mathcal{B}). \qquad (۱۲.۴)$$

از بخش ۱.۴ به‌خاطر بیاورید که Γ^n، بسط n ام Γ، یک کانال با الفبای ورودی و خروجی به‌ترتیب A^n و B^n است. ارسال هر کلمه $a_{i_۱} \dots a_{i_n} \in A^n$ از طریق Γ^n به‌وسیله ارسال متوالی سمبل‌های آن یعنی $a_{i_۱}$، ...، a_{i_n} از طریق Γ، و یا معادلاً با ارسال هم‌زمان آنها از طریق n کپی مستقل از Γ، صورت می‌گیرد. بنابراین احتمال‌های به‌جلو Γ^n به‌فرم زیر هستند

$$Pr(b_{i_۱} \dots b_{i_n}|a_{i_۱} \dots a_{i_n}) = Pr(b_{i_۱}|a_{i_۱}) \dots Pr(b_{i_n}|a_{i_n}).$$

اگر از $\mathcal{A}^n$ برای تعریف توزیع احتمال ورودی Γ^n استفاده کنیم، آنگاه توزیع خروجی به‌وسیله $\mathcal{B}^n$ تعیین می‌شود. از قضیه ۲۰.۳ نتیجه می‌شود:

$$H(\mathcal{A}^n) = nH(\mathcal{A}), \qquad H(\mathcal{B}^n) = nH(\mathcal{B}), \qquad (۱۳.۴)$$

و مشابهاً از تمرین ۴.۴ نتیجه می‌شود

$$\begin{aligned} H(\mathcal{A}^n|\mathcal{B}^n) &= nH(\mathcal{A}|\mathcal{B}), \\ H(\mathcal{B}^n|\mathcal{A}^n) &= nH(\mathcal{B}|\mathcal{A}), \\ H(\mathcal{A}^n, \mathcal{B}^n) &= nH(\mathcal{A}, \mathcal{B}). \end{aligned} \qquad (۱۴.۴)$$

اگر به‌عوض $\mathcal{A}$، ایده متوسط گیری کدهای شانون – فانو را روی $\mathcal{A}^n$ به‌کار ببریم آنگاه بنابر (۱۲.۴) یک کدگذاری برای $\mathcal{A}^n$ با متوسط طول کلمه L_n به‌دست می‌آوریم که در رابطه

$$H(\mathcal{A}^n|\mathcal{B}^n) \leq L_n \leq ۱ + H(\mathcal{A}^n|\mathcal{B}^n),$$

صدق کرده، و بنابراین براساس (۱۴.۴):

$$nH(\mathcal{A}|\mathcal{B}) \leq L_n \leq ۱ + nH(\mathcal{A}|\mathcal{B}).$$

به‌عنوان یک روش کدگذاری برای $\mathcal{A}$، این یکتا دکدپذیر بوده و دارای متوسط طول کلمه $\frac{L_n}{n}$ است. از تقسیم کردن بر n نتیجه می‌شود

$$H(\mathcal{A}|\mathcal{B}) \leq \frac{L_n}{n} \leq \frac{۱}{n} + H(\mathcal{A}|\mathcal{B});$$

بنابراین اگر $n \longrightarrow \infty$ آنگاه $\frac{L_n}{n} \longrightarrow H(\mathcal{A}|\mathcal{B})$.
این قضیه شانون را اثبات می‌کند که نظیر قضیه ۲۳.۳ می‌باشد:

قضیه ۸.۴ اگر خروجی $\mathcal{B}$ از یک کانال معلوم باشد آنگاه با انتخاب n به‌قدر کافی بزرگ و کدکردن $\mathcal{A}^n$ می‌توان یک کدگذاری یکتا دکدپذیر برای $\mathcal{A}$ ارائه داد که متوسط طول کلمه آن به‌قدر دلخواه به $H(\mathcal{A}|\mathcal{B})$ نزدیک باشد.

همچون حالت کدگذاری منبع، می‌توان نشان داد که متوسط طول کلمه هرگز کمتر از این کران نخواهد بود. قضایای ۸.۴ و ۲۳.۳ نشان می‌دهند که بسته به این‌که $\mathcal{B}$ معلوم یا غیرمعلوم باشد اعداد $H(\mathcal{A}|\mathcal{B})$ و $H(\mathcal{A})$ معرف اطلاعات حمل‌شده به‌وسیله $\mathcal{A}$، و نیز معرف کوچک‌ترین متوسط طول کلمه در کدگذاری $\mathcal{A}$ هستند. بنابراین از هر کانال انتظار می‌رود که در $H(\mathcal{A}|\mathcal{B}) \leq H(\mathcal{A})$ صدق کند، زیرا از $\mathcal{A}$ نمی‌توان اطلاعات بیشتری را وقتی $\mathcal{B}$ معلوم است در مقایسه با زمانی‌که $\mathcal{B}$ نامعلوم است به‌دست آورد. این نامساوی را برای BSC در گزاره (۱۰.۴) ثابت کردیم و در بخش بعد آن‌را در حالت کلی اثبات می‌کنیم.

۶.۴ اطلاعات متقابل

اگر Γ یک کانال با ورودی و خروجی $\mathcal{A}$ و $\mathcal{B}$ باشد آنگاه آنتروپی $\mathcal{A}$، $H(\mathcal{A})$، سه تفسیر معادل دارد:

۱) این آنتروپی معرف ابهام راجع به $\mathcal{A}$ است وقتی $\mathcal{B}$ نامعلوم است؛

۲) این آنتروپی میزان اطلاعات حمل‌شده به‌وسیله $\mathcal{A}$ است وقتی $\mathcal{B}$ نامعلوم است؛

۳) این آنتروپی می‌نیمم متوسط طول کلمه لازم برای کدکردن $\mathcal{A}$ است وقتی $\mathcal{B}$ معلوم نیست.

مشابهاً ابهام $H(\mathcal{A}|\mathcal{B})$ دارای سه تفسیر معادل است:

۱) این عدد معرف ابهام راجع به $\mathcal{A}$ است وقتی $\mathcal{B}$ معلوم است؛

۲) این عدد برابر اطلاعات حمل‌شده به‌وسیله $\mathcal{A}$ است وقتی $\mathcal{B}$ معلوم است؛

۳) این عدد برابر می‌نیمم متوسط طول کلمه لازم برای کدکردن $\mathcal{A}$ به شرط معلوم بودن $\mathcal{B}$ است.

تفاضل این دو عدد را به‌عنوان اطلاعات متقابل تعریف می‌کنیم

$$I(\mathcal{A},\mathcal{B}) = H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B}).$$

مشابه آنچه در بالا بیان شد، این عدد نیز سه تفسیر معادل دارد:

۱) این میزان ابهام مربوط به $\mathcal{A}$ است که با دانستن $\mathcal{B}$ برطرف می‌شود؛

۲) این میزان اطلاعات درباره $\mathcal{A}$ است که به‌وسیله $\mathcal{B}$ حمل می‌شود؛

۳) این عدد برابر می‌نیمم میانگین تعداد سمبل‌ها در کدکلمه‌های $\mathcal{A}$ است که به $\mathcal{B}$ مربوط می‌شوند.

تمامی این تعبیرات نشان می‌دهد که $I(\mathcal{A},\mathcal{B})$ معرف میزان اطلاعات مشترک $\mathcal{A}$ و $\mathcal{B}$ است. اگر به تشبیه با مجموعه‌های متناهی که در بخش ۳.۴ به‌کار رفت ادامه دهیم، می‌توان $I(\mathcal{A},\mathcal{B})$ را متناظر با فصل مشترک مجموعه‌ها درنظر گرفت، زیرا $|A\cap B| = |A| - |A\backslash B|$.

مثال ۹.۴ به‌عنوان یک مثال نسبتاً ساده، فرض کنید Γ یک شرکت فیلم بوده، $\mathcal{A}$ یک کتاب، و $\mathcal{B}$ فیلم حاصل از کتاب باشند. در این‌صورت $I(\mathcal{A},\mathcal{B})$ معرف میزان اطلاعاتی است که فیلم درباره کتاب به ما می‌گوید.

مثال ۱۰.۴ فرض کنید $\mathcal{A}$ یک درس، Γ یک دانشجو که یادداشت بر می‌دارد، و $\mathcal{B}$ مجموعه حاصل از یادداشت‌های درسی باشند. در این‌صورت $I(\mathcal{A},\mathcal{B})$ میزان دقت در ضبط نمودن درس را نشان می‌دهد.

از تعویض نقش $\mathcal{A}$ و $\mathcal{B}$ نتیجه می‌شود

$$I(\mathcal{B},\mathcal{A}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}),$$

که میزان اطلاعات آزاد شده راجع به $\mathcal{B}$ به‌وسیله $\mathcal{A}$ است (به‌عنوان نمونه، میزان اطلاعاتی که با خواندن کتاب نسبت به فیلم حاصل می‌شود). این مشابه رابطه $|B \cap A| = |B| - |B \backslash A|$ است. در (۶.۴) و (۷.۴) ملاحظه کردیم که

$$\begin{aligned} H(\mathcal{A},\mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}|\mathcal{A}), \\ H(\mathcal{A},\mathcal{B}) &= H(\mathcal{B}) + H(\mathcal{A}|\mathcal{B}). \end{aligned}$$

با حذف $H(\mathcal{A},\mathcal{B})$ از این دو معادله، می‌بینیم که

$$H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}),$$

بنابراین

$$I(\mathcal{A},\mathcal{B}) = I(\mathcal{B},\mathcal{A}). \qquad (۱۵.۴)$$

پس میزان اطلاعاتی که خروجی نسبت به ورودی در بردارد دقیقاً همان مقدار از اطلاعات است که ورودی نسبت به خروجی حمل می‌کند. اگر از (۷.۴) برای جایگزینی $H(\mathcal{A}|\mathcal{B})$ در تعریف $I(\mathcal{A},\mathcal{B})$ استفاده کنیم خواهیم داشت

$$I(\mathcal{A},\mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A},\mathcal{B}). \qquad (۱۶.۴)$$

(این مشابه $|A \cap B| = |A| + |B| - |A \cup B|$ است، همچنان‌که (۱۵.۴) متناظر با $|A \cap B| = |B \cap A|$ می‌باشد.)

قضیه ۱۱.۴ برای هر کانال Γ داریم $I(\mathcal{A},\mathcal{B}) \geq 0$، و تساوی برقرار است اگر و فقط اگر $\mathcal{A}$ و $\mathcal{B}$ از نظر آماری مستقل باشند.

اثبات از معادله (۱۶.۴) و به‌کار بردن $p_i = \sum_j R_{ij}$ و $q_j = \sum_i R_{ij}$ نتیجه می‌شود

$$\begin{aligned} I(\mathcal{A},\mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A},\mathcal{B}) \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} - \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_i \sum_j R_{ij} \log \frac{1}{p_i} + \sum_i \sum_j R_{ij} \log \frac{1}{q_j} - \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}} \\ &= \sum_i \sum_j R_{ij} \log \frac{1}{p_i q_j} - \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}}. \end{aligned}$$

چون $\sum_i \sum_j p_i q_j = 1$، پس می‌توان نتیجه ۹.۳ را برای توزیع‌های (R_{ij}) و $(p_i q_j)$ به‌کار برد. (به‌منظور برقرار نمودن شرایط نتیجه ۹.۳، می‌توان فرض کرد

که همواره $p_iq_j > 0$؛ این کار با ندیده گرفتن سمبل‌های ورودی و خروجی با احتمال صفر صورت می‌گیرد که در واقع تاثیری روی دستگاه آنتروپی ندارند.) نتیجه ۹.۳ نشان می‌دهد

$$\sum_i \sum_j R_{ij} \log \frac{1}{p_iq_j} \geq \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}},$$

که در آن تساوی برقرار است اگر و فقط اگر برای هر i و j داشته باشیم $R_{ij} = p_iq_j$، یعنی این‌که اگر و فقط اگر $\mathcal{A}$ و $\mathcal{B}$ مستقل باشند. ■

نتیجه ۱۲.۴ برای هر کانال Γ داریم

$$H(\mathcal{A}) \geq H(\mathcal{A}|\mathcal{B}), \quad H(\mathcal{B}) \geq H(\mathcal{B}|\mathcal{A}), \quad H(\mathcal{A},\mathcal{B}) \leq H(\mathcal{A}) + H(\mathcal{B});$$

در هر یک از حالات تساوی برقرار است اگر و فقط اگر $\mathcal{A}$ و $\mathcal{B}$ مستقل آماری باشند.

اثبات این حکم نتیجه فوری قضیه ۱۱.۴ و معادله زیر است

$$\begin{aligned} I(\mathcal{A},\mathcal{B}) &= H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B}) \\ &= H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}) \\ &= H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A},\mathcal{B}). \ \blacksquare \end{aligned}$$

توضیح مختصری از این نتایج را در بخش بعد ارائه می‌دهیم.

۷.۴ اطلاعات متقابل برای کانال دوتایی متقارن

به‌عنوان مثالی از محاسبه اطلاعات متقابل، فرض کنید Γ کانال BSC با نمادهای معمول بخش ۲.۴ باشد. در بخش ۶.۴ دیدیم که اطلاعات متقابل یک کانال عبارت است از

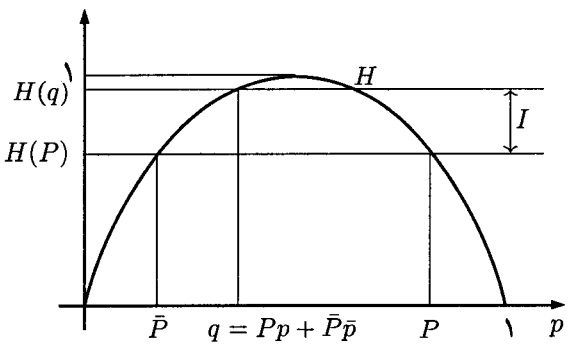
$$I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}).$$

در بخش ۴.۴ ملاحظه شد که BSC دارای خواص $H(\mathcal{B}) = H(q)$ و $H(\mathcal{B}|\mathcal{A}) = H(P)$ است که $q = pP + \bar{p}\overline{P}$، پس

$$\begin{aligned} I(\mathcal{A},\mathcal{B}) &= H(q) - H(P) \\ &= H(pP + \bar{p}\overline{P}) - H(P). \end{aligned}$$

توجه کنید که این به هر دو احتمال p و P، یعنی کانال و ورودی آن، بستگی دارد. گراف تابع H (شکل ۶.۴) نشان می‌دهد که $0 \leq I(\mathcal{A},\mathcal{B}) \leq 1 - H(P)$.

برای یک مقدار ثابت P، ماکزیمم مقدار برای $I(\mathcal{A},\mathcal{B})$ برابر $H(P)$ – ۱ است که متناظر با $p=\frac{1}{2}$ (یعنی $q=\frac{1}{2}$) بوده، و می‌نیمم مقدار آن ۰ است که متناظر با $p\in\{0,1\}$ می‌باشد. این اکسترمم‌ها متناظر با حالاتی هستند که سمبل‌های ورودی ثابت بوده و یا توزیع یکنواخت داشته باشند.



شکل ۶.۴

تمرین ۸.۴ فرض کنید $I=I(\mathcal{A},\mathcal{B})$ اطلاعات متقابل برای BSC باشد؛ در $\mathbf{R}^3$، رویه معرف I را به‌عنوان تابعی از p و P رسم کنید. این کار را برای BEC نیز انجام دهید.

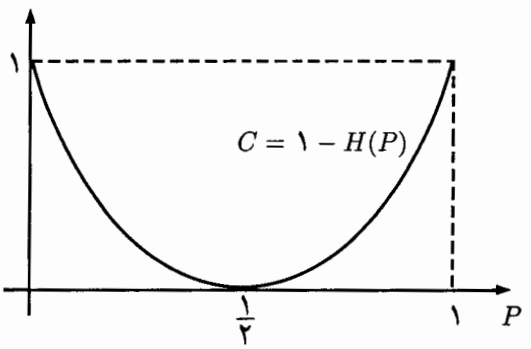
۸.۴ ظرفیت کانال

اطلاعات متقابل $I(\mathcal{A},\mathcal{B})$ برای یک کانال Γ معرف میزان اطلاعات راجع به ورودی $\mathcal{A}$ است که در خروجی $\mathcal{B}$ ظاهر می‌شود. همچنان‌که در بخش ۷.۴ برای حالت BSC ملاحظه کردیم، این بستگی به Γ و $\mathcal{A}$ دارد. برای یک کانال مفروض Γ، می‌خواهیم این مقدار را با انتخاب مناسب $\mathcal{A}$ (یا با انتخاب ورودی به‌عنوان کدگذاری مناسبی از یک منبع مفروض) ماکزیمم کنیم.

ظرفیت C از کانال Γ را برابر ماکزیمم اطلاعات متقابل $I(\mathcal{A},\mathcal{B})$ تعریف می‌کنیم که $\mathcal{A}$ روی تمامی ورودی‌های ممکن برای Γ تغییر می‌کند؛ بنابراین احتمال‌های به‌جلو P_{ij} ثابت بوده درحالی‌که احتمال‌های ورودی p_i متغیر می‌باشند. این بدین معنی است که C، که تنها به Γ بستگی دارد، معرف ماکزیمم مقدار اطلاعاتی است که کانال می‌تواند منتقل کند.

مثال ۱۳.۴ در انتهای بخش ۷.۴ دیدیم که کانال BSC دارای ظرفیت $C=1-H(P)$ است که با توزیع یکنواخت ورودی به‌دست می‌آید. شکل ۷.۴ ظرفیت را به‌عنوان تابعی از P نمایش می‌دهد. توجه کنید که C بیشترین مقدار خود را وقتی به‌دست می‌آورد که P برابر ۰ یا ۱ باشد، یعنی این‌که Γ کاملاً قابل اعتماد و یا کاملاً غیر قابل اعتماد (و بنابراین کاملاً

قابل پیش بینی) باشد؛ C کمترین مقدار خود را به‌ازای $P = \frac{1}{2}$، یعنی وقتی‌که Γ کمترین امکان پیش‌بینی را فراهم می‌کند، می‌گیرد.



شکل ۷.۴

تمرین ۹.۴ با به‌کار بردن تمرین ۷.۴، اطلاعات متقابل BEC را پیدا کرده و ظرفیت آن‌را تعیین کنید.

تمرین ۱۰.۴ نشان دهید که اگر کانال‌های Γ و Γ' ظرفیت‌های C و C' داشته باشند، آنگاه ضرب آنها $\Gamma \times \Gamma'$ دارای ظرفیت $C + C'$ بوده و ظرفیت Γ^n برابر nC است.

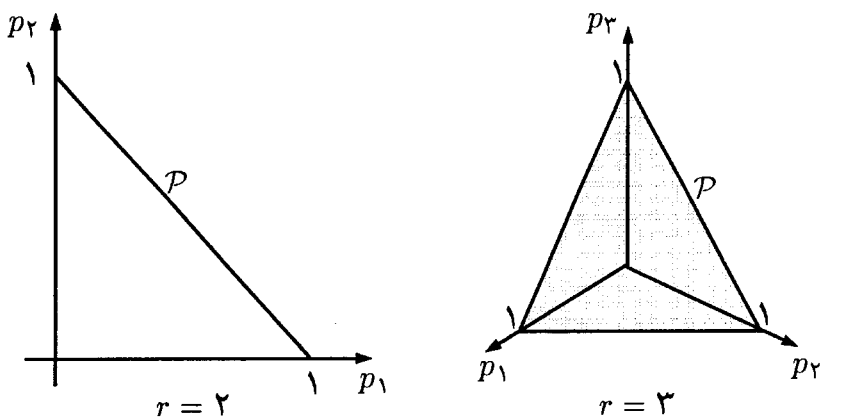
برای یک کانال دلخواه Γ، محاسبه C می‌تواند مشکل باشد. علت آن این است که این امر مستلزم ماکزیمم کردن تابع $I(\mathcal{A}, \mathcal{B})$ می‌باشد که غیرخطی بوده، برحسب لگاریتم احتمالاتی است که مقید به شرط $\sum p_i = ۱$ (یا معادلاً $\sum q_j = ۱$) می‌باشند. برای مثالی ساده در این رابطه به تمرین ۱۶.۴ مراجعه کنید.

برای تصدیق تعریفی که برای ظرفیت ارائه کردیم باید نشان دهیم که C همیشه وجود دارد، یعنی این‌که برای هر کانال Γ اطلاعات متقابل $I(\mathcal{A}, \mathcal{B})$ از بالا محدود بوده و کمترین کران بالای خود را اختیار می‌کند. (غیرمنطقی نیست که تصور شود $I(\mathcal{A}, \mathcal{B})$ محدود نبوده یا از بالا محدود باشد ولی کمترین کران بالای خود را اختیار نکند.) برای انجام این کار ایده‌ها و نتایجی از آنالیز را به‌کار می‌گیریم؛ خوانندگانی که نسبت به این موضوع حساسیت داشته و یا آمادگی قبول وجود C را دارند می‌توانند از بقیه این بخش صرف‌نظر کنند.

اجازه دهید Γ را ثابت گرفته و $\mathcal{A}$ را متغیر فرض کنیم، پس ماتریس کانال $M = (P_{ij})$ را ثابت گرفته و اجازه می‌دهیم که بردار توزیع احتمال ورودی $\mathbf{p} = (p_۱, \ldots, p_r)$ تغییر کند. دامنه تغییرات این بردار مجموعه $\mathcal{P}$ است:

$$\mathcal{P} = \{\mathbf{p} \in \mathbf{R}^r | p_i \geq ۰, \sum_i p_i = ۱\}.$$

از نظر هندسی $\mathcal{P}$ یک سیمپلکس $r-1$ بعدی است، مجموعه محدب محدود به r بردار مبنای استاندارد در $\mathbf{R}^r$. وقتی‌که $r=2$ یا $r=3$، این سیمپلکس به‌ترتیب یک پاره‌خط یا یک مثلث است (شکل ۸.۴)، و برای $r=4$ این مجموعه یک چهاروجهی در $\mathbf{R}^4$ است.



شکل ۸.۴

یک زیرمجموعه $X \subseteq \mathbf{R}^r$ بسته است هرگاه برای هر دنباله همگرای $\mathbf{x}_n \in X$، حد این دنباله $\lim_{n\to\infty} \mathbf{x}_n$ نیز در X باشد؛ معادلاً اگر $\mathbf{y} \notin X$ آنگاه هر نقطه به‌قدر کافی نزدیک به $\mathbf{y}$ نیز در خارج X قرار دارد. گوییم X محدود است هرگاه یک عدد حقیقی M موجود باشد به‌قسمی‌که برای هر $\mathbf{x} \in X$ داشته باشیم $|\mathbf{x}| \leq M$. یک زیرمجموعه بسته و محدود $X \subseteq \mathbf{R}^r$ را فشرده می‌نامیم[۲]. اثبات فشردگی $\mathcal{P}$ سرراست است (تمرین ۱۱.۴).

تمرین ۱۱.۴ نشان دهید که مجموعه $\mathcal{P}$ متشکل از بردارهای توزیع احتمال $\mathbf{p} = (p_1, \ldots, p_r)$ یک زیرمجموعه بسته و محدود از $\mathbf{R}^r$ است.

در بخش ۶.۴ دیدیم که $I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A})$. با توجه به رابطه

$$H(\mathcal{B}|\mathcal{A}) = \sum_i \sum_j p_i P_{ij} \log \frac{1}{P_{ij}} = \sum_i \left(\sum_j P_{ij} \log \frac{1}{P_{ij}} \right) p_i$$

و این‌که P_{ij} اعداد ثابتی هستند نتیجه می‌شود $H(\mathcal{B}|\mathcal{A})$ یک تابع خطی از $\mathbf{p}$، و بنابراین یک تابع پیوسته از $\mathbf{p}$ است. مشابهاً، چون $x \log x$ تابعی پیوسته از x روی $[0, 1]$ است، $H(\mathcal{B})$

$$H(\mathcal{B}) = \sum_j q_j \log \frac{1}{q_j}$$

[۲] در آنالیز قضیه هانیه — بورل نشان می‌دهد که این معادل با تعریف کلی‌تری از فشردگی برحسب مجموعه‌های باز است.

یک تابع پیوسته از $(q_j) = \mathbf{q}$ است. چون $\mathbf{q} = \mathbf{p}M$ و $M = (P_{ij})$ یک ماتریس ثابت است پس $H(\mathcal{B})$ یک تابع پیوسته از $\mathbf{p}$ است. در واقع $H(\mathcal{B})$ ترکیبی از دو تابع پیوسته است:

$$\mathbf{p} \longmapsto \mathbf{p}M = \mathbf{q} = (q_j) \longmapsto \sum_j q_j \log q_j = H(\mathcal{B}).$$

چون $I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A})$ و تفاضل دو تابع پیوسته تابعی پیوسته است پس قضیه بعدی را اثبات کرده‌ایم.

قضیه ۱۴.۴ اطلاعات متقابل $I(\mathcal{A},\mathcal{B})$ هر کانال Γ یک تابع پیوسته از بردار توزیع احتمال ورودی $\mathbf{p} = (p_i)$ است.

نتیجه ۱۵.۴ اطلاعات متقابل $I(\mathcal{A},\mathcal{B})$ هر کانال Γ دارای مقدار ماکزیمم است.

اثبات $I(\mathcal{A},\mathcal{B})$ یک تابع پیوسته از $\mathbf{p}$ بوده و دامنه $\mathbf{p}$ یک مجموعه فشرده $\mathcal{P}$ است. یک قضیه از آنالیز (به‌عنوان نمونه [La 83] را ببینید) می‌گوید که یک تابع حقیقی پیوسته روی یک مجموعه فشرده از بالا محدود بوده و کوچک‌ترین کران بالای خود را اختیار می‌کند، از این‌رو این مقدار ماکزیمم تابع است. ■

این تعریف ما از ظرفیت C از کانال Γ را تصدیق می‌کند. در فصل ۵ نشان خواهیم داد که C همچنین برابر پایین‌ترین کران بالا برای نسبت‌هایی است که با آن نسبت‌ها می‌توان اطلاعات را با دقت از طریق Γ ارسال نمود.

۹.۴ تمرینات تکمیلی

تمرین ۱۲.۴ فرض کنید، مانند تمرین ۱.۴، n کپی یکسان از کانال دوتایی متقارن Γ به‌صورت متوالی به‌هم متصل شده‌اند. نشان دهید که کانال حاصل نیز یک کانال دوتایی متقارن بوده و ظرفیت آن‌را محاسبه کنید. (راهنمایی: مقادیر ویژه ماتریس‌های کانال را درنظر بگیرید.) وضعیت را برای $n \to \infty$ توضیح دهید.

تمرین ۱۳.۴ نشان دهید که برای یک کانال Γ رابطه $C = 0$ برقرار است اگر و فقط اگر سطرهای ماتریس کانال یکسان باشند. تفسیر این وضعیت از دیدگاه دریافت کننده چگونه است؟

تمرین ۱۴.۴ تابع آنتروپی $H(\mathbf{p}) = -\sum_i p_i \log p_i$ را می‌توان به‌عنوان یک تابع $\mathbf{R} \longrightarrow \mathcal{P}$ درنظر گرفت که $\mathcal{P}$ مجموعه تمامی بردارهای توزیع احتمال $\mathbf{p} = (p_1,\ldots,p_r) \in \mathbf{R}^r$ است. نشان دهید که H روی $\mathcal{P}$ اکیداً محدب است، به این معنی که اگر $\mathbf{p},\mathbf{q} \in \mathcal{P}$ و $\lambda \in [0,1]$ آنگاه $H(\lambda\mathbf{p} + \bar{\lambda}\mathbf{q}) \geq \lambda H(\mathbf{p}) + \bar{\lambda}H(\mathbf{q})$ که در آن تساوی برقرار است اگر و فقط اگر $\mathbf{p} = \mathbf{q}$ یا $\lambda \in \{0,1\}$.

تمرین ۱۵.۴ یک کانال Γ یکنواخت است هرگاه هر سطر و هر ستون ماتریس کانال جایگشتی از سطر اول باشد. نشان دهید ظرفیت Γ برابر

$$\log s + \sum_{j=1}^{s} P_{ij} \log P_{ij},$$

است که به‌وسیله توزیع یکنواخت ورودی حاصل می‌شود. از این‌جا ظرفیت کانال r–تایی متقارن را پیدا کنید که در آن $r = s$، $P_{ii} = P$، و $P_{ij} = \overline{P}/(r-1)$ اگر $i \neq j$.

تمرین ۱۶.۴ یک کانال دوتایی کلی Γ با یک ماتریس (P_{ij}) مشخص می‌شود که در آن $P_{i1} + P_{i2} = 1$ برای $i = 1, 2$. نشان دهید که اطلاعات متقابل Γ برابر است با

$$I(\mathcal{A}, \mathcal{B}) = -q_1 \log q_1 - q_2 \log q_2 + q_1 c_1 + q_2 c_2,$$

که q_1 و q_2 احتمال‌های خروجی بوده و c_1 و c_2 به‌قسمی انتخاب می‌شوند که برای $i = 1, 2$ تساوی $P_{i1}c_1 + P_{i2}c_2 = P_{i1} \log P_{i1} + P_{i2} \log P_{i2}$ برقرار باشد. نتیجه بگیرید که Γ دارای ظرفیت $C = \log(2^{c_1} + 2^{c_2})$ است. (راهنمایی: از روش ضرایب لاگرانژ برای ماکزیمم کردن $I(\mathcal{A}, \mathcal{B})$ مشروط به $q_1 + q_2 = 1$ استفاده کنید.) چه می‌شود اگر $P_{11} = P_{22}$؟ (این تمرین بر مبنای کار مورگا [Mu 53] است.)

تمرین ۱۷.۴ نشان دهید اگر Γ_1 و Γ_2 دارای ظرفیت‌های C_1 و C_2 باشند آنگاه مجموع آنها یعنی $\Gamma_1 + \Gamma_2$ دارای ظرفیت $\log(2^{C_1} + 2^{C_2})$ است. این نتیجه را برای حالت $\Gamma_1 = \Gamma_2$ چگونه تفسیر می‌کنید؟

تمرین ۱۸.۴ فرض کنید Γ یک آبشار $\Gamma_1 \circ \Gamma_2$ از کانال‌های Γ_1 و Γ_2 باشد که Γ_1 ورودی $\mathcal{A}$ و خروجی $\mathcal{B}$، و Γ_2 ورودی $\mathcal{B}$ و خروجی $\mathcal{C}$ دارد. نشان دهید

$$H(\mathcal{A}|\mathcal{C}) - H(\mathcal{A}|\mathcal{B}) = \sum_b \sum_c (Pr(b,c) \sum_a Pr(a|b)(\log Pr(a|b) - \log Pr(a|c))),$$

که جمع‌ها روی سمبل‌های a، b و c از $\mathcal{A}$ و $\mathcal{B}$ و $\mathcal{C}$ صورت می‌گیرند. نتیجه بگیرید $H(\mathcal{A}|\mathcal{C}) \geq H(\mathcal{A}|\mathcal{B})$، و یک توضیح شهودی درباره این نتیجه ارائه کنید. از این‌جا قضیه پردازش داده‌ها، $I(\mathcal{A}, \mathcal{C}) \leq I(\mathcal{A}, \mathcal{B})$، را اثبات کنید که نشان می‌دهد اطلاعات متقابل را با ارسال بیشتر نمی‌توان افزایش داد (بحث مشابه‌ای نشان می‌دهد $I(\mathcal{A}, \mathcal{C}) \leq I(\mathcal{B}, \mathcal{C})$). نشان دهید اگر هر Γ_i ظرفیت C_i داشته باشد آنگاه Γ دارای ظرفیت $C \leq \min(C_1, C_2)$ است؛ مثال‌هایی ارائه دهید که $C = \min(C_1, C_2)$ و $C < \min(C_1, C_2)$.

فصل ۵

به‌کارگیری یک کانال غیر قابل اعتماد

اجازه دهید به یک چنین مردی اعتماد نشود.(بازرگان ونیزی)

در این فصل فرض می‌کنیم که Γ یک کانال غیر قابل اعتماد، مانند یک BSC با $P < ۱$، بوده و وظیفه ما انتقال اطلاعات از طریق Γ با بیشترین دقت ممکن است. قضیه اساسی شانون[۱] که احتمالاً مهمترین نتیجه در نظریه اطلاعات است می‌گوید ظرفیت C، از کانال Γ، کوچک‌ترین کران بالا برای نرخ‌هایی است که با آنها می‌توان اطلاعات را سالم از طریق Γ ارسال نمود. بعد از توضیح چند مفهوم درگیر در بحث، مثال ساده‌ای را در رابطه با این‌که چگونه می‌توان به این انتقال دقیق دست یافت ملاحظه خواهیم کرد. یک اثبات کامل برای قضیه شانون از نظر فنی کاملاً دشوار است، از این‌رو، برای سادگی بحث، ما اثبات را به حالت BSC محدود می‌کنیم؛ برای این کانال یک اثبات حاوی نکات اصلی را در بخش ۴.۵، و سپس اثبات کامل را در ضمیمه C خواهید یافت.

۱.۵ قواعد تصمیم

فرض کنید Γ یک کانال اطلاعات با ورودی $\mathcal{A}$ و خروجی $\mathcal{B}$ باشد. دریافت کننده که هر سمبل خروجی $b = b_j$ خارج شده از کانال Γ را می‌بیند نیاز به الگوریتمی دارد تا تصمیم

[۱] Shannon's fundamental theorem

بگیرد که کدام‌یک از سمبل‌های ورودی $a = a_i \in A$ منجر به b_j شده است. این سبب یک قاعده تصمیم، یعنی یک تابع $\Delta : B \longrightarrow A$ می‌شود. هر وقت b_j از Γ ظاهر شود دریافت کننده تابع Δ را روی b_j به‌کار برده و با تعیین $a_i = \Delta(b_j)$ تصمیم می‌گیرد (احتمالاً نادرست) که a_i ارسال شده است؛ این عمل را کدگشایی (Decoding) خروجی می‌نامیم. با نوشتن $i = j^*$ خواهیم داشت $\Delta(b_j) = a_{j^*}$.

مشکل این است که در حالت کلی توابع بسیاری چون $\Delta : B \longrightarrow A$ وجود دارند و خیلی واضح نیست که کدام‌یک برای استفاده بهترین است.

تمرین ۱.۵ چند قاعده تصمیم متمایز برای یک کانال اطلاعات مفروض موجود است؟

مثال ۱.۵ فرض کنید Γ کانال BSC باشد، پس $A = B = \mathbf{Z}_2$. اگر دریافت کننده به این کانال اعتماد کند آنگاه Δ باید تابع همانی باشد، یعنی $\Delta(0) = 0$ و $\Delta(1) = 1$؛ در غیر این‌صورت تابع دیگری $\Delta : \mathbf{Z}_2 \longrightarrow \mathbf{Z}_2$ باید به‌کار رود (مثال ۵.۴ را ملاحظه کنید که مربوط به وضعیتی می‌شود که در آن انتخاب $\Delta(0) = \Delta(1) = 0$ منطقی است).

اگر b_j دریافت شود، آنگاه دریافت کننده a_{j^*} را به‌عنوان سمبل ارسال شده درنظر می‌گیرد. احتمال درست بودن این تصمیم برابر است با

$$Pr(a = a_{j^*} | b = b_j) = Q_{j^*j}.$$

(بخش ۱.۴ را برای تعاریف و نمادهای احتمال‌هایی که در این‌جا به‌کار می‌روند ببینید.) هر b_j با احتمال q_j دریافت می‌شود، پس با متوسط‌گیری روی $b_j \in B$ می‌بینیم که متوسط احتمال کدگشایی درست، Pr_C، برابر است با

$$Pr_C = \sum_j q_j Q_{j^*j} = \sum_j R_{j^*j}, \qquad (۱.۵)$$

که $R_{ij} = q_j Q_{ij}$ احتمال توأم $Pr(a_i, b_j)$ است. نتیجه می‌شود که احتمال خطا (متوسط احتمال کدگشایی نادرست) برابر است با

$$Pr_E = 1 - Pr_C = 1 - \sum_j R_{j^*j} = \sum_j \sum_{i \neq j^*} R_{ij}. \qquad (۲.۵)$$

برای کانال و ورودی مفروض Γ و $\mathcal{A}$، می‌خواهیم قاعده تصمیمی چون $\Delta : B \longrightarrow A$ انتخاب کنیم که Pr_E را می‌نیمم کند (معادلاً Pr_C را ماکزیمم کند)؛ یک چنین قاعده‌ای بعضاً مشاهده‌گر ایده‌آل نامیده می‌شود. برای هر j، مقدار $i = j^*$ را آنچنان انتخاب می‌کنیم که احتمال به‌عقب $Pr(a_i | b_j) = Q_{ij}$ ماکزیمم شود. (اگر چند i در این شرط صدق کنند

آنگاه یکی از آنها را به‌دلخواه به‌عنوان j^* انتخاب می‌کنیم.) این معادل با ماکزیمم کردن احتمال توأم $R_{ij} = q_j Q_{ij}$ برای هر j است، یعنی این‌که $R_{j^*j} \geq R_{ij}$ برای هر i؛ پس R_{j^*j} بزرگ‌ترین عضو در ستون j از ماتریس (R_{ij}) است. چون $R_{ij} = p_i P_{ij}$، این ماتریس را می‌توان از ماتریس کانال $M = (P_{ij})$ و توزیع ورودی (p_i) به‌شکل زیر به‌دست آورد که‌در آن درایه‌های جاهای خالی همگی صفر هستند

$$(R_{ij}) = \begin{pmatrix} p_1 & & \\ & \ddots & \\ & & p_r \end{pmatrix} M.$$

مثال ۲.۵ اگر Γ کانال BSC باشد آنگاه بنابر بخش ۲.۴ داریم

$$(R_{ij}) = \begin{pmatrix} p & 0 \\ 0 & \overline{p} \end{pmatrix} \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix} = \begin{pmatrix} pP & p\overline{P} \\ \overline{p}\,\overline{P} & \overline{p}P \end{pmatrix}.$$

پس Δ به‌فرم زیر تعریف می‌شود

$$\Delta(0) = \begin{cases} 0 & pP > \overline{p}\,\overline{P} \\ 1 & pP < \overline{p}\,\overline{P}, \end{cases} \qquad \Delta(1) = \begin{cases} 1 & \overline{p}P > p\overline{P} \\ 0 & \overline{p}P < p\overline{P}, \end{cases}$$

که در حالت تساوی مقدار $\Delta(b)$ را دلخواه تعریف می‌کنیم.

تمرین ۲.۵ مقدار Pr_E را برای کانال Γ و ورودی A معرفی شده در مثال ۵.۴ محاسبه کنید (یک BSC با $P = 0.8$ و $p = 0.9$)، که Δ مشاهده‌گر ایده‌آل است.

در بعضی از شرایط، ممکن است گیرنده رفتار کانال را بداند ولی از ورودی مطلع نباشد، به‌طوری‌که احتمال‌های به‌جلو P_{ij} معلوم بوده ولی توزیع ورودی p_i مشخص نباشد. این بدین معنی است که احتمال‌های Q_{ij} و R_{ij} معلوم نبوده و نمی‌توان از آنها در انتخاب قاعده تصمیم Δ استفاده کرد. در چنین وضعیتی، دریافت کننده باید انتخاب Δ را تنها بر مبنای P_{ij} که تنها به کانال بستگی دارند انجام دهد. روش بدیهی این است که برای هر j عدد $i = j^*$ را به‌قسمی انتخاب کنیم که P_{ij} را ماکزیمم کند، از این‌رو P_{j^*j} بزرگ‌ترین درایه در ستون j از ماتریس $M = (P_{ij})$ است. (طبق معمول اگر چند درایه با این خاصیت موجود باشند یکی از آنها را به دلخواه انتخاب می‌کنیم.) قاعده به‌دست آمده Δ که بر مبنای آن برای هر i رابطه $P_{j^*j} \geq P_{ij}$ برقرار است قاعده بیشترین درستنمایی نامیده می‌شود؛ همچون قبل، احتمال‌های Pr_C و Pr_E با روابط (۱.۵) و (۲.۵) مشخص می‌شوند. اگر سمبل‌های ورودی هم‌احتمال باشند، آنگاه برای هر j احتمال‌های به‌جلو P_{ij} متناسب با احتمال‌های توأم هستند $R_{ij} = p_i P_{ij} = P_{ij}/r$، پس در این حالت قاعده بیشترین درستنمایی با قاعده

مشاهده‌گر ایده‌آل یکی می‌شود. برای توزیع‌های ورودی دیگر ممکن است این قاعده بهترین نباشد (مثال ۴.۵ را ببینید)؛ در هر حال، در میان تمامی توزیع‌ها این توزیع از جهت ماکزیمم نمودن انتگرال چند گانه

$$\int_{\mathbf{p}\in\mathcal{P}} Pr_C dp_1 \dots dp_r$$

بهترین است که در آن $\mathcal{P}$ مجموعه متشکل از بردارهای توزیع احتمال $\mathbf{p} = (p_1, \dots, p_r) \in \mathbf{R}^r$ است. براساس این نتیجه درک طبیعی شخص درست است: اگر اطلاعی از ورودی در دست نباشد آنگاه توزیع‌های مختلف هم‌وزن بوده و قاعده بیشترین درستنمایی بهترین قاعده خواهد بود.

تمرین ۳.۵ ادعای بالا را ثابت کنید که می‌گوید در میان تمامی قواعد تصمیم برای یک کانال مفروض، قاعده بیشترین درستنمایی انتگرال Pr_C را روی تمامی ورودی‌های $\mathbf{p} \in \mathcal{P}$ ماکزیمم می‌کند.

مثال ۳.۵ اجازه دهید قاعده بیشترین درستنمایی Δ را روی کانال BSC به‌کار ببریم که در آن $P > \frac{1}{2}$ (پس Γ بیشتر قابل اعتماد است تا غیر قابل اعتماد). پس $P > \overline{P}$، از این‌رو با انتخاب بزرگ‌ترین درایه در هر ستون از ماتریس کانال

$$M = \begin{pmatrix} P & \overline{P} \\ \overline{P} & P \end{pmatrix}$$

تعریف می‌کنیم $\Delta(0) = 0$ و $\Delta(1) = 1$. نتیجه می‌شود

$$Pr_C = pP + \overline{p}P = P, \qquad Pr_E = p\overline{P} + \overline{p}\,\overline{P} = \overline{P}.$$

از طرف دیگر، اگر $P < \frac{1}{2}$ آنگاه $\overline{P} > P$، و از این‌رو $\Delta(1) = 0$ و $\Delta(0) = 1$. بنابراین

$$Pr_C = \overline{p}\overline{P} + p\overline{P} = \overline{P}, \qquad Pr_E = \overline{p}P + pP = P.$$

مثال ۴.۵ برای یک توضیح مشخص، به مثال ۵.۴ باز می‌گردیم که در آن $P = 0.8$ و $p = 0.9$. همچنان‌که در مثال ۳.۵ دیدیم از قاعده بیشترین درستنمایی نتیجه می‌شود $\Delta(0) = 0$ و $\Delta(1) = 1$، با $Pr_C = P = 0.8$. ولی از قاعده مشاهده‌گر ایده‌آل نتیجه می‌شود $\Delta(0) = \Delta(1) = 0$، با $Pr_C = 0.9 > 0.8$ (تمرین ۲.۵ را ببینید)؛ پس در این‌جا قاعده بیشترین درستنمایی بهترین انتخاب نیست.

مثال ۵.۵ فرض کنید Γ کانال BEC مثال ۲.۴ با $P > ۰$ باشد. در این‌صورت از قاعده بیشترین درستنمایی داریم $\Delta(۰) = ۰$، $\Delta(۱) = ۱$، و $\Delta(?)$ می‌تواند ۰ یا ۱ باشد، مثلاً $\Delta(?) = ۰$. پس اگر احتمال‌های ورودی برای ۰ و ۱ برابر p و $\overline{p}$ باشند آنگاه

$$Pr_C = pP + \overline{p}P + p\overline{P} = P + p\overline{P}, \qquad Pr_E = \overline{p}.۰ + \overline{p}.۰ + \overline{p}\,\overline{P} = \overline{p}\,\overline{P}.$$

۲.۵ مثالی از اعتماد بهبود یافته

از طریق یک کانال غیر قابل اعتماد مفروض، چگونه می‌توان اطلاعات را با اعتماد بیشتر ارسال کرد؟ قبل از این‌که این مسئله را در حالت کلی درنظر بگیریم اجازه دهید به یک مثال ساده بپردازیم. Γ را کانال BSC با $\frac{۱}{۲} < P < ۱$ درنظر می‌گیریم؛ برای سهولت در نمادگذاری تعریف می‌کنیم $Q = \overline{P} = ۱ - P$، از این‌رو ماتریس کانال به‌شکل زیر است

$$M = \begin{pmatrix} P & Q \\ Q & P \end{pmatrix}.$$

چون $P > Q$، مثال ۳.۵ نشان می‌دهد که قاعده بیشترین درستنمایی با $\Delta(۰) = ۰$، $\Delta(۱) = ۱$، و $Pr_E = Q$ مشخص می‌شود. همچنین فرض می‌کنیم سمبل‌های ورودی هم‌احتمال باشند، یعنی $p = \overline{p} = \frac{۱}{۲}$؛ پس اطلاعات متقابل $I(\mathcal{A}, \mathcal{B})$ بیشترین مقدار خود را که ظرفیت کانال، $C = ۱ - H(P)$، است اختیار می‌کند (بخش‌های ۷.۴ و ۸.۴ را ببینید).

اگر احتمال خطا $Pr_E = Q$ به‌شکل غیر قابل قبولی بالا باشد آنگاه برای کاهش دادن این احتمال متناظر با هر سمبل ورودی $a \in \{۰, ۱\}$، دنباله aaa را ارسال می‌کنیم. این بدین معنی است که از کد

$$\mathcal{C} : ۰ \longmapsto ۰۰۰, \quad ۱ \longmapsto ۱۱۱,$$

استفاده می‌کنیم. بنابراین ورودی $\mathcal{C}$ متشکل از دو کلمه هم‌احتمال $w = ۰۰۰$ و $w = ۱۱۱$ است. در حین ارسال w از طریق Γ، هر یک از سه سمبل کلمه w می‌تواند تغییر یابد، بنابراین خروجی $\mathcal{D}$ شامل هشت کلمه دوتایی به‌طول سه است. اکنون هر سمبل از کلمه w با احتمال P یا Q به‌ترتیب درست یا نادرست ارسال می‌شود، از این‌رو احتماهای به‌جلو برای ورودی و خروجی جدید با ماتریس زیر مشخص می‌شود

$$\begin{pmatrix} P^۳ & P^۲Q & P^۲Q & P^۲Q & PQ^۲ & PQ^۲ & PQ^۲ & Q^۳ \\ Q^۳ & PQ^۲ & PQ^۲ & PQ^۲ & P^۲Q & P^۲Q & P^۲Q & P^۳ \end{pmatrix},$$

که سطرها از بالا به پایین متناظر با ۰۰۰ و ۱۱۱ بوده و ستون j ام متناظر با دنباله دوتایی نمایش $۱ - j$ است. چون $P > Q > ۰$ داریم $P^۳ > Q^۳$ و $P^۲Q > PQ^۲$، پس قاعده

بیشترین درستنمایی چنین است

$$\Delta : \begin{cases} ۰۰۰, ۰۰۱, ۰۱۰, ۱۰۰ \longmapsto ۰۰۰, \\ ۰۱۱, ۱۰۱, ۱۱۰, ۱۱۱ \longmapsto ۱۱۱. \end{cases}$$

با ترکیب نمودن این با معکوس $\mathcal{C}$

$$۰۰۰ \longmapsto ۰, \quad ۱۱۱ \longmapsto ۱,$$

می‌توان کلمات $\mathcal{D}$ را با قاعده زیر کدگشایی کرد

$$۰۰۰, ۰۰۱, ۰۱۰, ۱۰۰ \longmapsto ۰,$$
$$۰۱۱, ۱۰۱, ۱۱۰, ۱۱۱ \longmapsto ۱.$$

این قاعده کدگشایی اکثریت نیز نامیده می‌شود: تعداد سمبل‌های ۰ و ۱ دریافت شده را شمرده و سمبل با بیشترین تکرار را به‌عنوان سمبل ارسال شده لحاظ می‌کنیم. با به‌کار بردن کلمات به‌طول فرد امکان انتخاب سمبل با بیشترین تکرار تضمین می‌شود.

$$\begin{matrix} ۰ \\ ۱ \end{matrix} \longrightarrow \begin{matrix} ۰۰۰ \\ ۱۱۱ \end{matrix} \longrightarrow \Gamma \longrightarrow \begin{matrix} ۰۰۰ \\ ۰۰۱ \\ ۰۱۰ \\ ۱۰۰ \\ ۰۱۱ \\ ۱۰۱ \\ ۱۱۰ \\ ۱۱۱ \end{matrix} \longrightarrow \begin{matrix} ۰ \\ ۱ \end{matrix}$$

شکل ۱.۵

تمامی روند کدگذاری، ارسال و کدگشایی در شکل ۱.۵ خلاصه شده است. در عمل یک کانال دوتایی متقارن جدید Γ' ساخته‌ایم؛ یک سمبل ورودی $a \in \{۰, ۱\}$ توسط $\mathcal{C}$ به کلمه $w \in \{۰۰۰, ۱۱۱\}$ کدگذاری شده و از طریق Γ ارسال می‌شود؛ کلمه دریافت شده با قاعده اکثریت به $b = ۰$ یا $b = ۱$ کدگشایی می‌شود. اکنون عمل کدگشایی درست است $(b = a)$ اگر و فقط اگر در حین ارسال w حداکثر یکی از سمبل‌های آن تغییر کرده باشد. هر سمبل w با احتمال P درست ارسال می‌شود، پس احتمال عدم وجود خطا در ارسال w برابر $P^۳$ است. برای وجود یک خطا سه حالت امکان‌پذیر است که در هر حالت یکی از سمبل‌ها تغییر یافته و دو سمبل دیگر درست دریافت می‌شوند؛ هر یک از این حالت‌ها احتمال $P^۲Q$ دارد، پس احتمال وجود یک خطا برابر $۳P^۲Q$ است. مشابهاً احتمال دو و سه خطا به‌ترتیب برابر $۳PQ^۲$ و $Q^۳$ است. پس ماتریس کانال برای Γ' برابر M' است

$$M' = \begin{pmatrix} P^۳ + ۳P^۲Q & ۳PQ^۲ + Q^۳ \\ ۳PQ^۲ + Q^۳ & P^۳ + ۳P^۲Q \end{pmatrix}.$$

پس Γ' یک BSC با احتمال ارسال درست $P' = P^۳ + ۳P^۲Q$ است. بنابراین احتمال خطا برابر است با

$$Pr_E = \overline{P'} = ۳PQ^۲ + Q^۳ = Q^۲(۳ - ۲Q) \approx ۳Q^۲,$$

که به‌طور قابل ملاحظه‌ای کمتر از احتمال خطای اولیه Q، برای مقدار کوچک $Q > ۰$، می‌باشد. (به‌عنوان نمونه، اگر $Q = ۰.۰۱$ آنگاه $Pr_E = ۰.۰۰۰۲۹۸$.) بنابراین احتمال خطا را به قیمت کندی ارسال بهبود بخشیده‌ایم: اکنون برای ارسال یک سمبل باید سه سمبل ارسال شود، از این‌رو می‌گوئیم که نرخ ارسال برابر $R = ۱/۳$ (در مقایسه با مقدار اولیه ۱) است.

تعمیم ساده‌ای از این ایده بدین شرح است که برای ارسال سمبل ۰ یک دنباله به‌طول n از ۰ و برای ارسال ۱ یک دنباله به‌طول n از ۱ ارسال شود؛ این کد تکراری دوتایی به‌طول n است که با $\mathcal{R}_n$ نمایش داده می‌شود. اگر n را فرد بگیریم آنگاه قاعده بیشترین درستنمایی کدگشایی اکثریت است، همچنان‌که در تمرین ۱۰.۵ نشان داده شده است. (اگر n زوج باشد، کلمه دریافت شده می‌تواند حاوی تعداد یکسانی از ۰ و ۱ بوده و بنابراین اکثریتی را ایجاد نکند.) می‌توان نشان داد که اگر $n \longrightarrow \infty$ آنگاه $Pr_E \longrightarrow ۰$ (تمرین ۱۰.۵). با فرض $Q = ۰.۰۱$، جدول زیر مقادیر تقریبی Pr_E را برای n های فرد کمتر از ۱۲ ارائه می‌دهد:

$n =$	۱	۳	۵	۷	۹	۱۱
$Pr_E \approx$	$۱۰^{-۲}$	۳×۱۰^{-۴}	$۱۰^{-۵}$	۳.۵×۱۰^{-۷}	۱.۳×۱۰^{-۸}	۵×۱۰^{-۱۰}

با این وجود، نرخ ارسال نیز به ۰ میل می‌کند $R = ۱/n \longrightarrow ۰$، بنابراین افزایش دقت را به قیمت کندی ارسال فراهم کرده‌ایم.

این ایده را می‌توان بیشتر تعمیم داد. اگر Γ یک کانال با ورودی A باشد که الفبای آن A دارای r سمبل است، آنگاه هر زیرمجموعه $\mathcal{C} \subseteq A^n$ را می‌توان به‌عنوان یک مجموعه از کدکلمه‌ها به‌کار برد که از طریق Γ ارسال می‌شوند. به‌عنوان نمونه، کد تکراری $\mathcal{R}_n$ روی A متشکل از تمامی کلمات به‌طول n به‌فرم $w = aa\ldots a$ است که $a \in A$. به چند مثال بیشتر در فصل‌های ۶ و ۷ خواهیم پرداخت. $\mathcal{C}$ را یک کد r–تایی به‌طول n می‌نامیم. اگر $|\mathcal{C}| = r^k$ آنگاه $\mathcal{C}$ می‌تواند A^k را کدگذاری کند (زیرا این منبع حاوی r^k کلمه است). در هر کدکلمه از $\mathcal{C}$، n سمبل معرف k سمبل ظاهر شده از A می‌باشد، از این‌رو می‌گوئیم نرخ این کد $R = k/n$ است. به‌عنوان مثال، کد تکراری r–تایی $\mathcal{R}_n$ دارای پارامترهای $|\mathcal{R}_n| = r$، $k = ۱$، و $R = ۱/n$ است. به‌شکل کلی‌تری، نرخ (یا نرخ ارسال) یک کد غیرخالی $\mathcal{C} \subseteq A^n$ برابر R تعریف می‌شود

$$R = \frac{\log_r |\mathcal{C}|}{n}, \qquad (۳.۵)$$

به‌طوری‌که $|\mathcal{C}| = r^{Rn}$. چون $|A^n| = r^n$ می‌بینیم که برای هر کد $\mathcal{C}$ رابطه $0 \leq R \leq 1$ برقرار است.

قضیه اساسی شانون (بخش ۴.۵) می‌گوید که با انتخاب کدهای $\mathcal{C} \subseteq A^n$، با درنظر گرفتن n به‌قدر کافی بزرگ، و با انتخاب قواعد تصمیم مناسب، می‌توان احتمال خطای Pr_E را به‌قدر دلخواه به ۰ نزدیک نمود بدون این‌که نرخ ارسال نیز به ۰ نزدیک شود (آنچنان‌که برای $\mathcal{R}_n$ به ۰ میل می‌کند). در واقع این عمل را می‌توان با نرخ R که به‌قدر دلخواه به ظرفیت کانال نزدیک باشد انجام داد.

۳.۵ فاصله همینگ

در بخش قبل با مثالی ساده چگونگی ارسال اطلاعات با دقت بیشتر تشریح شد. یکی از مؤلفه‌های مهم روش ارائه شده انتخاب کدکلمه‌های ۰۰۰۰۰ و ۱۱۰۰۱ است که خیلی با هم تفاوت دارند، به‌قسمی‌که حتی اگر با خطا دریافت شوند دریافت کننده هنوز احتمالاً قادر به تشخیص آنها خواهد بود. وقتی بخواهیم این ایده را برای ساخت کدهای مؤثرتری به‌کار ببریم نیاز به انتخاب مجموعه‌های بزرگ‌تری از کدکلمه‌ها داریم که با هم تفاوت زیادی نیز داشته باشند. برای اندازه‌گیری شباهت یا تمایز بین دو کلمه، تعریفی از فاصله بین دو کلمه را معرفی می‌کنیم.

فرض کنید $\mathbf{u} = u_1 \cdots u_n$ و $\mathbf{v} = v_1 \cdots v_n$ دو کلمه به‌طول n روی یک الفبای A باشند، پس $\mathbf{v}, \mathbf{u} \in A^n$. (این کلمات را به‌شکل پررنگ می‌نویسم زیرا نیاز به این خواهیم داشت که آنها را به‌عنوان بردار در یک فضای برداری A^n درنظر بگیریم که A یک میدان است.) فاصله همینگ[1] بین $\mathbf{u}$ و $\mathbf{v}$ را با $d(\mathbf{u}, \mathbf{v})$ نمایش داده و برابر تعداد اندیس‌های i تعریف می‌کنیم که $u_i \neq v_i$.

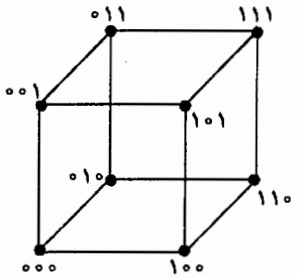
مثال ۶.۵ فرض کنید در $\mathbf{Z}_2^5$ داشته باشیم $\mathbf{u} = ۰۱۱۰۱$ و $\mathbf{v} = ۰۱۰۰۰$. آنگاه $d(\mathbf{u}, \mathbf{v}) = ۲$، زیرا کلمات $\mathbf{u}$ و $\mathbf{v}$ در دو موقعیت $(i = ۳, ۵)$ با هم متفاوت‌اند.

تمرین ۴.۵ اگر $\mathbf{u} \in A^n$ که $|A| = r$، و $0 \leq i \leq n$، آنگاه چند کلمه $\mathbf{v} \in A^n$ با خاصیت $d(\mathbf{u}, \mathbf{v}) = i$ وجود دارد؟ نشان دهید که مجموعه این اعداد برای $i = ۰, ۱, \ldots, n$ برابر $|A^n|$ است.

مثال ۷.۵ می‌توان کلمات در $\mathbf{Z}_2^3$ را به‌عنوان ۸ رأس یک مکعب درنظر گرفت (شکل ۲.۵).

در این‌صورت $d(\mathbf{u}, \mathbf{v})$ فاصله اقلیدسی نیست، بلکه کمترین تعداد اضلاع در مسیرهای بین $\mathbf{u}$ و $\mathbf{v}$ است. این تعریف از فاصله در نظریه گراف به‌کار می‌رود که در آن فاصله بین دو

[1] Hamming distance



شکل ۲.۵

رأس در یک گراف همبند برابر طول (تعداد ضلع) کوتاه‌ترین مسیر بین آن دو رأس تعریف می‌شود.

تمرین ۵.۵ اگر برای هر دو عضو متمایز $\mathbf{u}$ و $\mathbf{v}$ از مجموعه $\mathcal{C} \subseteq \mathbf{Z}_2^3$ داشته باشیم $d(\mathbf{u},\mathbf{v}) \geq 2$، آنگاه بزرگی اندازه $\mathcal{C}$ تا چه حدی می‌تواند باشد؟ یک توصیف هندسی برای تمامی مجموعه‌هایی که دارای این حد هستند ارائه دهید. مشابه این کران را در $\mathbf{Z}_2^n$ مشخص کنید.

لم ۸.۵ فرض کنید $\mathbf{w},\mathbf{u},\mathbf{v} \in A^n$. در این‌صورت

(الف) $d(\mathbf{u},\mathbf{v}) \geq 0$، و $d(\mathbf{u},\mathbf{v}) = 0$ اگر و فقط اگر $\mathbf{u} = \mathbf{v}$؛

(ب) $d(\mathbf{u},\mathbf{v}) = d(\mathbf{v},\mathbf{u})$؛

(ج) $d(\mathbf{u},\mathbf{w}) \leq d(\mathbf{u},\mathbf{v}) + d(\mathbf{v},\mathbf{w})$.

اثبات قسمت‌های (الف) و (ب) واضح بوده و قسمت (ج) نیز آسان است (تمرین ۶.۵).■

قسمت (ج) به نامساوی مثلث معروف است، زیرا اشاره به این واقعیت دارد که یک ضلع $\mathbf{uw}$ از مثلث $\mathbf{uvw}$ نمی‌تواند طولی بزرگ‌تر از مجموع طول دو ضلع دیگر $\mathbf{uv}$ و $\mathbf{vw}$ داشته باشد.

تمرین ۶.۵ بخش (ج) لم ۸.۵ را ثابت کنید.

در توپولوژی، یک مجموعه با یک تابع d واجد خواص سه‌گانه بیان شده را به‌عنوان یک فضای متریک می‌شناسند، اگرچه ما این واقعیت را به‌کار نخواهیم برد. نکته مورد نظر این نتیجه این است که نشان دهد رفتار d خیلی شبیه به تابع فاصله اقلیدسی در $\mathbf{R}^n$ است.

برای ارسال اطلاعات از طریق Γ، یک کد $\mathcal{C} \subseteq A^n$ انتخاب کرده و قاعده بیشترین درستنمایی را به‌کار می‌بریم، یعنی این‌که هر کلمه دریافتی را به کدکلمه‌ای دکد می‌کنیم که با

بیشترین احتمال باعث دریافت آن کلمه شده باشد. این بهترین تصمیم خواهد بود اگر کدکلمه‌ها توزیع یکنواخت داشته باشند، و در حالت کلی بهترین است اگر احتمال ورودی معلوم نباشد. حتی اگر برای بعضی توزیع احتمال خاص، این قاعده تصمیم‌گیری بهترین نباشد ولی آن‌قدر خوب هست تابتواند در اثبات قضیه شانون برای n به‌قدر کافی بزرگ مقدار Pr_E را به‌قدر دلخواه به ۰ نزدیک کند.

برای سادگی فرض خواهیم کرد که در بقیه این بخش و بخش بعدی Γ کانال BSC با $P > \frac{1}{2}$ باشد، پس $A = B = \mathbf{Z}_2$ و $r = 2$. انتخاب ما از قاعده بیشترین درستنمایی به این معنی است که برای هر خروجی $\mathbf{v} \in \mathbf{Z}_2^n$، ما $\mathbf{v}$ را به کدکلمه $\mathbf{u} = \Delta(\mathbf{v}) \in \mathcal{C}$ که ماکزیمم کننده احتمال به‌جلو $Pr(\mathbf{v}|\mathbf{u})$ باشد دکد می‌کنیم. حال اگر $d(\mathbf{u}, \mathbf{v}) = i$ آنگاه

$$Pr(\mathbf{v}|\mathbf{u}) = Q^i P^{n-i},$$

بنابراین

$$Pr(\mathbf{v}|\mathbf{u}) = P^n (Q/P)^i,$$

که یک تابع نزولی از i است زیرا $Q/P < 1$. پس یک کدکلمه $\mathbf{u}$ احتمال $Pr(\mathbf{v}|\mathbf{u})$ را ماکزیمم می‌کند اگر و فقط اگر $d(\mathbf{u}, \mathbf{v})$ را می‌نیمم کند. پس قاعده بیشترین درستنمایی Δ هر کلمه دریافتی $\mathbf{v} \in \mathbf{Z}_2^n$ را به کدکلمه‌ای چون $\mathbf{u} = \Delta(\mathbf{v}) \in \mathcal{C}$ دکد می‌کند که نسبت به فاصله همینگ نزدیک‌ترین کدکلمه به $\mathbf{v}$ باشد. این قاعده، که در بقیه فصل ۵ از آن استفاده خواهیم کرد، کدگشایی نزدیک‌ترین همسایه نامیده می‌شود. طبق معمول، اگر بیش از یک نزدیک‌ترین همسایه موجود باشد یکی از آنها را به‌دلخواه انتخاب می‌کنیم.

۴.۵ بیان و رئوس اثبات قضیه شانون

نتیجه زیر که اغلب قضیه اساسی نظریه اطلاعات نامیده می‌شود در ۱۹۴۸ به‌وسیله شانون اثبات شد [Sh 48]. به‌صورت غیر رسمی، این قضیه می‌گوید اگر از کدکلمه‌های به‌قدر کافی طولانی استفاده کنیم می‌توان اطلاعات را از طریق کانال Γ با دقت دلخواه و با نرخ به‌قدر دلخواه نزدیک به ظرفیت کانال ارسال کرد. این یک بهبود در مقایسه با مثال کد تکراری $\mathcal{R}_n$ است که در آن دقت مورد نیاز با انتخاب n به‌قدر کافی بزرگ و بنابراین با میل دادن نرخ ارسال به سمت ۰، و نه ظرفیت کانال، به‌دست می‌آید. اگرچه قضیه برای تمامی کانال‌ها معتبر است ولی برای سادگی، آن‌را برای کانال BSC بیان و ثابت می‌کنیم. بیان دقیق قضیه به‌شرح زیر است.

قضیه ۹.۵ فرض کنید Γ یک کانال دوتایی متقارن با $P > \frac{1}{2}$ باشد، بنابراین ظرفیت کانال برابر $C = 1 - H(P) > 0$ است. نیز فرض کنید δ و ε دو عدد مثبت باشند. در این‌صورت

برای تمامی اعداد طبیعی به‌قدر کافی بزرگ n، یک کد $\mathcal{C} \subseteq \mathbf{Z}_2^n$ وجود دارد که نرخ آن R در شرط $C - \varepsilon \leq R < C$ صدق کرده، و کدگشایی نزدیک‌ترین همسایه منجر به $Pr_E < \delta$ می‌شود.

پس با انتخاب δ و ε به‌قدر کافی کوچک، می‌توانیم Pr_E و R را به‌ترتیب به‌قدر دلخواه به ۰ و C نزدیک کنیم. یک اثبات کامل را که نسبتاً طولانی است در ضمیمه C خواهید دید. در این‌جا تنها ایده‌های اصلی اثبات را می‌آوریم.

رئوس اثبات فرض کنید $R < C$ (به‌قدر دلخواه نزدیک)، و برای عدد بزرگ n به‌صورت تصادفی یک مجموعه $\mathcal{C}$ شامل 2^{nR} کلمه از $\mathbf{Z}_2^n$ انتخاب کنید. (از این‌که ممکن است 2^{nR} عدد صحیح نباشد صرف‌نظر می‌کنیم! در یک اثبات دقیق، یک عدد صحیح نزدیک به 2^{nR} انتخاب کرده و نشان می‌دهیم که این تعدیل نتیجه را تحت تأثیر قرار نخواهد داد.) این یک کد دوتایی $\mathcal{C}$ به‌طول n ایجاد کرده که بنابر (۳.۵) نرخ آن $\frac{1}{n}\log_2 2^{nR} = R$ است.

اگر یک کدکلمه $\mathbf{u} \in \mathcal{C}$ از طریق Γ ارسال شود، آنگاه احتمال خطای هر یک از n سمبل کلمه $\mathbf{u}$ برابر $\overline{P} = Q$ بوده، و بنابراین انتظار این را داریم که تعداد nQ از سمبل‌ها نادرست ارسال شوند. در واقع قانون اعداد بزرگ (ضمیمه B) ایجاب می‌کند که با $n \longrightarrow \infty$ این وضع با احتمال متمایل به ۱ رخ دهد. این بدین معنی است که باید انتظار این‌را داشته باشیم که کلمه دریافتی $\mathbf{v}$ در شرط $d(\mathbf{u}, \mathbf{v}) \approx nQ$ صدق کند. معادلاً از دید دریافت کننده، اگر کلمه $\mathbf{v}$ دریافت شود آنگاه کدکلمه ارسالی در شرط $d(\mathbf{u}, \mathbf{v}) \approx nQ$ صدق می‌کند.

قاعده نزدیک‌ترین همسایه هر کلمه دریافتی $\mathbf{v}$ را به کدکلمه $\Delta(\mathbf{v}) \in \mathcal{C}$ که نزدیک‌ترین کدکلمه به $\mathbf{v}$ است دکد می‌کند؛ پس اگر کدگشایی نادرست باشد آنگاه باید کدکلمه‌ای چون $\mathbf{u}' \neq \mathbf{u}$ در $\mathcal{C}$ با خاصیت $d(\mathbf{u}', \mathbf{v}) \leq d(\mathbf{u}, \mathbf{v})$ موجود باشد. از این‌رو احتمال کدگشایی نادرست $\mathbf{v}$ بیشتر از احتمال وجود کدکلمه‌ای چون $\mathbf{u}'$، با خاصیت یادشده، نیست. بنابراین

$$Pr_E \leq \sum_{\mathbf{u}' \neq \mathbf{u}} Pr(d(\mathbf{u}', \mathbf{v}) \leq nQ). \qquad (۴.۵)$$

در (۴.۵) عدد $d(\mathbf{u}, \mathbf{v})$ را با مقدار تقریبی آن یعنی nQ جایگزین نموده و از تغییر کمی که از این طریق در احتمال ایجاد می‌شود صرف‌نظر کرده‌ایم. چون تعداد $|\mathcal{C}| - 1 = 2^{nR} - 1$ کدکلمه $\mathbf{u}' \neq \mathbf{u}$ وجود دارد و به‌صورت تصادفی انتخاب می‌شوند، در (۴.۵) کران بالا روی Pr_E برابر است با

$$(|\mathcal{C}| - 1)Pr(d(\mathbf{u}', \mathbf{v}) \leq nQ) < 2^{nR} Pr(d(\mathbf{u}', \mathbf{v}) \leq nQ).$$

حال چون کدکلمه $\mathbf{u}'$ را به‌شکل تصادفی از $\mathbf{Z}_۲^n$ انتخاب کردیم، برای هر $\mathbf{v}$ مفروض احتمال وقوع نامساوی $d(\mathbf{u}', \mathbf{v}) \leq nQ$ برابر است با درصد ۲^n کلمه $\mathbf{u}' \in \mathbf{Z}_۲^n$ که در این نامساوی صدق می‌کنند. برای هر $\mathbf{v}$ و i مفروض، تعداد کلمات $\mathbf{u}' \in \mathbf{Z}_۲^n$ با خاصیت $d(\mathbf{u}', \mathbf{v}) = i$ برابر ضریب دوجمله‌ای $\binom{n}{i}$ است. بنابراین تعداد کلمات $\mathbf{u}' \in \mathbf{Z}_۲^n$ با خاصیت $d(\mathbf{u}', \mathbf{v}) \leq nQ$ برابر است با $\sum_{i \leq nQ} \binom{n}{i}$. پس

$$Pr(d(\mathbf{u}', \mathbf{v}) \leq nQ) = \frac{۱}{۲^n} \sum_{i \leq nQ} \binom{n}{i}.$$

برای ادامه اثبات نیاز به نتیجه زیر داریم که در اثبات کامل قضیه شانون نیز به‌کار می‌رود.

تمرین ۷.۵ نشان دهید اگر $۰ \leq \lambda \leq \frac{۱}{۲}$ و $\lambda + \mu = ۱$، آنگاه

$$۱ \geq \sum_{i \leq \lambda n} \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i \leq \lambda n} \binom{n}{i} \lambda^{\lambda n} \mu^{\mu n},$$

و از این‌رو نشان دهید

$$\sum_{i \leq \lambda n} \binom{n}{i} \leq ۲^{nH(\lambda)}.$$

(این‌را با اتّحاد معروف $\sum_{i=۰}^{n} \binom{n}{i} = ۲^n$ مقایسه کنید.)

ادامه اثبات حال به اثبات قضیه ۹.۵ باز می‌گردیم. اگر در تمرین ۷.۵ قرار دهیم $\lambda = Q$، آنگاه

$$\sum_{i \leq nQ} \binom{n}{i} \leq ۲^{nH(Q)},$$

پس (۴.۵) تبدیل به نامساوی زیر می‌شود

$$Pr_E < ۲^{nR} . \frac{۱}{۲^n} . ۲^{nH(Q)} = ۲^{n(R-۱+H(Q))} = ۲^{n(R-C)},$$

زیرا ظرفیت Γ برابر $C = ۱ - H(P) = ۱ - H(Q)$ است (بخش ۸.۴). حال چون $R < C$، اگر $n \longrightarrow \infty$ آنگاه $۲^{n(R-C)} \longrightarrow ۰$، و بنابراین $Pr_E \longrightarrow ۰$. ■

تذکر احتمالاً متوجه وجود چند شکاف در این اثبات شده‌اید و ممکن است شکاف‌های دیگری نیز وجود داشته باشند که شما متوجه نشده باشید، مگر اثبات را با دقت مطالعه کرده باشید. به‌عنوان نمونه، در جمله پایانی باید مطمئن شویم که $R - C$ با $n \longrightarrow \infty$ از ۰ فاصله داشته، به‌قسمی‌که $n(R - C) \longrightarrow -\infty$. با این وجود، می‌توان یک اثبات دقیق بر مبنای طرح بالا ارائه داد. چون فصل‌های بعدی وابسته به این اثبات نیستند آن‌را در بخش ضمیمه C قرار داده‌ایم.

۵.۵ عکس قضیه شانون

قضیه شانون می‌گوید که می‌توان اطلاعات را از طریق Γ، با دقت دلخواه، و با نرخ $R < C$ که به‌قدر دلخواه به ظرفیت C نزدیک باشد ارسال کرد. یک سوال واضح این است که پرسیده شود آیا بهتر از این هم می‌توان عمل کرد، یعنی این‌که آیا می‌توان در این‌جا ظرفیت C را با عدد ثابت دیگری چون C' جایگزین کرد که $C' > C$. در این بخش نشان می‌دهیم که این امکان‌پذیر نیست، و از این‌رو C کوچک‌ترین کران بالای نرخ‌هایی است که با آنها ارسال با دقت دلخواه امکان‌پذیر است. این را نه‌تنها برای کانال BSC بلکه برای یک کانال دلخواه ثابت می‌کنیم. ابتدا به نامساوی فانو[۱] نیاز داریم که یک کران پایین برای احتمال خطا ارائه می‌دهد.

قضیه ۱۰.۵ فرض کنید Γ یک کانال با ورودی $\mathcal{A}$ و خروجی $\mathcal{B}$ باشد. برای هر تابع تصمیم Δ برای Γ، احتمال خطا Pr_E در نامساوی زیر صدق می‌کند که r تعداد سمبل‌های $\mathcal{A}$ است:

$$H(\mathcal{A}|\mathcal{B}) \leq H(Pr_E) + Pr_E \log(r - 1). \qquad (۵.۵)$$

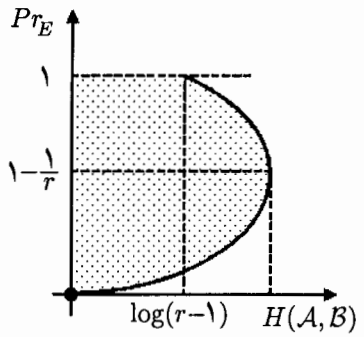
مقادیری از Pr_E و $H(\mathcal{A}|\mathcal{B})$ که در این نامساوی صدق می‌کنند با ناحیه سایه خورده در شکل ۳.۵ مشخص شده‌اند. قبل از اثبات قضیه، اجازه دهید این نتیجه را تفسیر کنیم. گیرنده، که سمبل خروجی b_j را می‌داند، قاعده تصمیم Δ را برای پیدا کردن $a_{j^*} = \Delta(b_j)$ به‌کار می‌برد، که این سمبل می‌تواند سمبل واقعی ارسال شده a_i باشد و می‌تواند هم نباشد. سمت چپ (۵.۵) ایهام Γ، یعنی میانگین اطلاعات اضافی که گیرنده برای دانستن a_i نیاز دارد، می‌باشد. این اطلاعات اضافی را می‌توان به دو بخش تقسیم کرد:

(الف) این‌که آیا کدگشایی درست است یا نه، یعنی این‌که آیا $a_{j^*} = a_i$ یا نه؛

(ب) اگر کدگشایی درست نیست، دراین‌صورت کدام $a_i (i \neq j^*)$ سمبل ارسال شده است.

نادرست و درست بودن عمل کدگشایی به‌ترتیب با احتمال‌های Pr_E و $Pr_C = \overline{Pr_E}$ است، و بنابراین مقدار اطلاعات بخش (الف) برابر $H(Pr_E)$ است که جمله اول سمت راست نامساوی (۵.۵) می‌باشد. اگر کدگشایی درست باشد گیرنده به هیچ اطلاعات دیگری نیاز ندارد، ولی اگر نادرست باشد (که با احتمال Pr_E اتفاق می‌افتد) گیرنده نیاز به این دارد که بداند کدام‌یک از $r - 1$ سمبل $a_i \neq a_{j^*}$ ارسال شده است. این منجر به انتخاب یک سمبل از میان $r - 1$ سمبل می‌شود، و بنابراین با قرار دادن $q = r - 1$ در قضیه ۱۰.۳ می‌بینیم که این اطلاعات حداکثر $\log(r - 1)$ است. با ضرب کردن این عدد در احتمال

[۱] Fano's inequality



شکل ۳.۵

این‌که (ب) لازم شود، ملاحظه می‌کنیم که به‌طور متوسط اطلاعات بخش (ب) حداکثر $Pr_E \log(r-۱)$ است که جمله دوم در سمت راست (۵.۵) می‌باشد.

اثبات قضیه ۱۰.۵ بدون از دست دادن کلیت مسئله می‌توانیم لگاریتم و آنتروپی را در پایه r محاسبه کنیم. در بخش ۳.۴ دیدیم که

$$H(\mathcal{A}|\mathcal{B}) = \sum_i \sum_j R_{ij} \log \frac{۱}{Q_{ij}} = -\sum_j \sum_i R_{ij} \log Q_{ij}.$$

جملات این جمع مضاعف را به دو مجموعه تقسیم می‌کنیم که متناظر با کدگشایی درست و کدگشایی نادرست (به‌ترتیب $i = j^*$ و $i \neq j^*$) هستند:

$$H(\mathcal{A}|\mathcal{B}) = -\sum_j R_{j^* j} \log Q_{j^* j} - \sum_j \sum_{i \neq j^*} R_{ij} \log Q_{ij}.$$

اکنون

$$H(Pr_E) = -Pr_E \log Pr_E - Pr_C \log Pr_C,$$

که

$$Pr_C = \sum_j R_{j^* j}, \qquad Pr_E = \sum_j \sum_{i \neq j^*} R_{ij}.$$

پس

$$H(\mathcal{A}|\mathcal{B}) - H(Pr_E) - Pr_E \log(r-۱) = \sum_j R_{j^* j} \log \frac{Pr_C}{Q_{j^* j}} + \sum_j \sum_{i \neq j^*} R_{ij} \log \frac{Pr_E}{(r-۱)Q_{ij}}.$$

از لم ۸.۳، با درنظر گرفتن مبنای r، برای هر $x > 0$ رابطه $\log x \leq (x-1)$ برقرار است. بنابراین برای سمت راست تساوی فوق داریم

$$\begin{aligned}
&\leq \log e \left(\sum_j R_{j^\bullet j} \left(\frac{Pr_C}{Q_{j^\bullet j}} - 1 \right) + \sum_j \sum_{i \neq j^\bullet} R_{ij} \left(\frac{Pr_E}{(r-1)Q_{ij}} - 1 \right) \right) \\
&= \log e \left(Pr_C \sum_j q_j - \sum_j R_{j^\bullet j} + \frac{Pr_E}{(r-1)} \sum_j \sum_{i \neq j^\bullet} q_j - \sum_j \sum_{i \neq j^\bullet} R_{ij} \right) \\
&= \log e \left(Pr_C - Pr_C + \frac{Pr_E}{(r-1)}(r-1) - Pr_E \right) \\
&= 0,
\end{aligned}$$

که در خط دوم از $R_{ij} = q_j Q_{ij}$، و در خط سوم از $\sum_j \sum_i q_j = r$ استفاده کرده‌ایم. ■

اکنون می‌توانیم چیزی را که به عکس قضیه شانون معروف است ثابت کنیم، یعنی این‌که نشان دهیم اگر $C' > C$ آنگاه این درست نیست که برای هر $\varepsilon > 0$ دنباله‌ای از کدهای $\mathcal{C}$، به‌طول $n \longrightarrow \infty$، و با نرخ R با خاصیت $C' - \varepsilon \leq R < C'$، وجود داشته باشد به‌قسمی‌که $Pr_E \longrightarrow 0$ وقتی $n \longrightarrow \infty$. برای اثبات این مطلب، کافی است نشان داده شود که برای مقداری از ε یک چنین دنباله‌ای وجود ندارد. اجازه دهید قرار دهیم $\varepsilon = (C' - C)/2$ و فرض کنیم یک چنین دنباله‌ای از کدهای $\mathcal{C}$ وجود دارد. چون $C' - \varepsilon \leq R < C'$ پس هر کد دارای نرخ $R \geq C + \varepsilon$ است.

می‌توانیم $\mathcal{C}$ را به‌عنوان ورودی برای Γ^n، بسط n ام Γ، با خروجی $\mathcal{D} = \mathcal{B}^n$ درنظر بگیریم. با به‌کار بردن کران فانو (قضیه ۱۰.۵) روی کانال Γ^n می‌بینیم

$$H(\mathcal{C}|\mathcal{D}) \leq H(Pr_E) + Pr_E \log(M-1),$$

که $M = |\mathcal{C}|$. معادلاً، می‌توانیم $\mathcal{C}$ را برای تعریف یک توزیع احتمال روی الفبای ورودی A^n درنظر بگیریم؛ هر یک از $M = |\mathcal{C}|$ کدکلمه در $\mathcal{C}$ احتمال $1/M$ داشته، و بقیه کلمات A^n با احتمال 0 لحاظ می‌شوند. بنابر تمرین ۱۰.۴، کانال Γ^n دارای ظرفیت nC بوده و از این‌رو

$$H(\mathcal{C}) - H(\mathcal{C}|\mathcal{D}) = I(\mathcal{C}, \mathcal{D}) \leq nC.$$

بنابراین از $M = r^{nR}$ نتیجه می‌شود

$$\begin{aligned}
H(\mathcal{C}) - nC &\leq H(\mathcal{C}|\mathcal{D}) \\
&\leq H(Pr_E) + Pr_E \log(M-1)
\end{aligned}$$

$$\begin{aligned} &\leq H(Pr_E) + Pr_E \log M \\ &= H(Pr_E) + Pr_E.nR. \end{aligned}$$

چون $\mathcal{C}$ دارای M کدکلمه با احتمال یکسان است، پس طبق قضیه ۱۰.۳ داریم $H(\mathcal{C}) = \log M = nR$ و در نتیجه

$$nR - nC \leq H(Pr_E) + Pr_E.nR.$$

بنابراین برای هر n داریم

$$0 < \varepsilon \leq R - C \leq \frac{1}{n}H(Pr_E) + Pr_E.R \leq \frac{1}{n}H(Pr_E) + Pr_E.$$

بااین وجود، اگر $Pr_E \longrightarrow 0$ آنگاه $H(Pr_E) \longrightarrow 0$، و از این‌رو عبارت سمت راست برای تمامی n های به‌قدر کافی بزرگ کوچک‌تر از ε است. پس $Pr_E \not\longrightarrow 0$. (در واقع، یک بحث دقیق‌تر نشان می‌دهد $Pr_E \longrightarrow 1$.)

مثال ۱۱.۵ فرض کنید Γ کانال BSC بوده، و به‌عنوان مثالی غیر واقعی از یک کد اجازه دهید قرار دهیم $\mathcal{C} = \mathcal{A}^n$، پس $R = 1$. اگر $0 < P < 1$ آنگاه $C = 1 - H(P) < 1$، و از این‌رو $R > C$. با به‌کار بردن تابع همانی $\Delta(\mathbf{u}) = \mathbf{u}$ به‌عنوان قاعده تصمیم، می‌بینیم که کدگشایی درست است اگر و فقط اگر خطایی وجود نداشته باشد، پس با $n \longrightarrow \infty$ خواهیم داشت $Pr_E = 1 - P^n \longrightarrow 1$.

مثال ۱۲.۵ در بخش ۴.۷ دنباله‌ای از کدهای دوتایی $\mathcal{H}_n$ (کدهای همینگ) به‌طول n، که n به‌شکل $2^c - 1$ است، و نرخ $R = (n-c)/n$ خواهیم ساخت. پس $R \longrightarrow 1$ هرگاه $n \longrightarrow \infty$. اگر از BSC با $0 < P < 1$ استفاده کنیم، آنگاه $C = 1 - H(P) < 1$ و بنابراین برای n به‌قدر کافی بزرگ داریم $R > C$. در بخش ۴.۷ خواهیم دید که کدگشایی نزدیک‌ترین همسایه درست است اگر و فقط اگر حداکثر یک خطا وجود داشته باشد؛ پس اگر $n \longrightarrow \infty$ آنگاه $Pr_E = 1 - P^n - nP^{n-1}Q$ به ۱ میل خواهد کرد.

۶.۵ نکاتی راجع به قضیه شانون

شکل کلی قضیه شانون به شرح زیر است:

قضیه ۱۳.۵ فرض کنید Γ یک کانال اطلاعات با ظرفیت $C > 0$ باشد، و فرض کنید $\delta, \varepsilon > 0$. برای تمامی n های به‌قدر کافی بزرگ یک کد $\mathcal{C}$ به‌طول n و نرخ R، با خاصیت $C - \varepsilon \leq R < C$، همراه با یک قاعده تصمیم وجود دارد که دارای احتمال خطای $Pr_E < \delta$ است.

اصول اساسی اثبات مشابه حالت BSC است؛ برای جزییات کامل [As 65] را ببینید. اگرچه این یک نتیجه خیلی قوی است با این حال چند محدودیت دارد:

نکته ۱۴.۵ برای رسیدن به نرخ R نزدیک به C و احتمال خطای Pr_E نزدیک به ۰، ممکن است مجبور به استفاده از مقادیر خیلی بزرگ n باشیم. این بدین معنی است که کدکلمه‌ها خیلی طولانی بوده و از این‌رو ممکن است اعمال کدگذاری و کدگشایی خیلی دشوار و زمان‌بر باشند. افزون بر این، اگر n بزرگ باشد آنگاه گیرنده به جهت منتظر ماندن برای دریافت یک کلمه کامل با تأخیر عمل می‌کند؛ وقتی یک کلمه دریافت شده دکد می‌شود یک حجم زیادی از اطلاعات به‌صورت لحظه‌ای ایجاد می‌شود که ممکن است رسیدگی به آن دشوار باشد.

نکته ۱۵.۵ قضیه شانون می‌گوید که کدهای خوب وجود دارند، ولی نه صورت قضیه و نه اثبات آن کمک زیادی به پیدا کردن این کدها می‌کنند. اثبات نشان می‌دهد که کد 'میانگین' خوب است، ولی هیچ تضمینی برای خوب بودن یک کد خاص وجود ندارد: این نیاز به بررسی همه جانبه آن کد دارد. اگر کسی یک کد به‌صورت تصادفی انتخاب کند، همچنان‌که در اثبات قضیه صورت گرفت، شانس معقولی برای خوب بودن آن وجود دارد. با این حال، کار کردن با کدهای تصادفی مشکل است. به‌صورت ایده‌آلی انتظار این است که یک کد از نظر ساختاری غنی باشد تا براساس آن بتوان الگوریتم‌هایی برای کدگذاری و کدگشایی طراحی کرد. مثال‌هایی از این نوع را در فصل‌های ٦ و ٧ به هنگام ساخت کدهای خاصی با نرخ ارسال و یا احتمال خطای خوب ملاحظه خواهید کرد.

۷.۵ تمرینات تکمیلی

تمرین ۸.۵ فرض کنید Γ کانال BEC با $P > ۰$ باشد، و فرض کنید احتمال‌های ورودی p و $\overline{p}$ باشند که $۰ < p < ۱$. نشان دهید چگونه باید از کد تکراری دوتایی $\mathcal{R}_n$ برای ارسال اطلاعات از طریق Γ استفاده کرد به‌قسمی‌که اگر $n \longrightarrow \infty$ آنگاه $Pr_E \longrightarrow ۰$.

تمرین ۹.۵ یک کانال دوتایی Γ سمبل ۰ را همیشه درست ارسال کرده، ولی ۱ را به‌عنوان ۱ و ۰ به‌ترتیب با احتمال‌های P و $Q = \overline{P}$ ارسال می‌کند که $۰ < P < ۱$. ماتریس کانال را نوشته و قاعده بیشترین درستنمایی را شرح دهید. اگر احتمال سمبل‌های ورودی ۰ و ۱ به‌ترتیب p و $\overline{p}$ باشند، مقدار Pr_E را تعیین کنید. برای افزایش اعتماد، سمبل‌های ۰ و ۱ به‌ترتیب با ۰۰۰ و ۱۱۱ کدگذاری می‌شوند. قاعده بیشترین درستنمایی حاصل را توضیح دهید؛ آیا این قاعده مشابه کدگشایی اکثریت است؟ در رابطه با تشابه آن با کدگشایی نزدیک‌ترین همسایه چطور؟ نرخ و احتمال خطای کد حاصل را پیدا کنید. اگر از کد تکراری دوتایی $\mathcal{R}_n$ استفاده کرده و اجازه دهیم n به بینهایت میل کند چه اتفاق می افتد؟

تمرین ۱۰.۵ کد دوتایی تکراری $\mathcal{R}_n$، به‌طول $n = ۲t + ۱$، برای ارسال اطلاعات از طریق کانال BSC Γ، که در آن هر بیت با احتمال P درست و با احتمال $Q = \overline{P}$ نادرست ارسال می‌شود و $P > \frac{۱}{۲}$، به‌کار می‌رود. نشان دهید که در این حالت قاعده بیشترین درستنمایی، کدگشایی اکثریت، و کدگشایی نزدیک‌ترین همسایه، همگی قاعده تصمیم مشابه‌ای چون Δ ارائه می‌دهند. نشان دهید احتمال خطای این قاعده در رابطه

$$Pr_E \leq \frac{(۲t+۱)!}{(t!)^۲} P^t Q^{t+۱},$$

صدق کرده، و نتیجه بگیرید اگر $n \longrightarrow \infty$ آنگاه $Pr_E \longrightarrow ۰$. چرا از این یک اثبات مستقیم برای قضیه اساسی شانون حاصل نمی‌شود؟

تمرین ۱۱.۵ (این تمرین و تمرین بعدی بر پایه مرجع [Ke 56] هستند.) یک قمارباز روی نتایج دنباله‌ای از پرتاب‌های یک سکه سالم شرط می‌بندد؛ شرط خود را بعد از پرتاب شدن سکه، ولی قبل از اعلام پیشامد رخ داده، مشخص می‌کند. اگر شرط را ببرد دو برابر گرو را دریافت و در صورت باخت به‌اندازه گرو می‌پردازد. او تصمیم به تقلب به‌وسیله مطلع شدن از نتیجه پرتاب از طریق یک کانال BSC Γ با احتمال‌های P و Q، به‌ترتیب برای ارسال درست و نادرست، می‌گیرد. بر این اساس روی سمبل ظاهر شده از Γ به میزان λ درصد سرمایه خود شرط بسته و روی سمبل دیگر مابقی سرمایه را، $\mu = \bar{\lambda}$، که λ ثابت است. نشان دهید که اگر سرمایه اولیه او $c_۰$ باشد، آنگاه بعد از n پرتاب برابر $c_n = ۲^n \lambda^m \mu^{n-m} c_۰$ است که m تعداد دفعاتی است که Γ اطلاعات درست می‌دهد. نشان دهید که در یک پریود طولانی از زمان نرخ رشد نمایی سرمایه قمارباز

$$G = \lim_{n \longrightarrow \infty} \frac{۱}{n} \log \left(\frac{c_n}{c_۰} \right)$$

احتمالاً با $\lambda = P$ این $G \approx ۱ + P \log \lambda + Q \log \mu$ مشخص می‌شود. نشان دهید با انتخاب مقدار ماکزیمم می‌شود که در این‌صورت $G \approx C$، که C ظرفیت Γ است. اگر $\frac{۱}{۲} < P < ۱$، قمارباز چگونه می‌تواند از مطالعه این فصل استفاده ببرد؟

تمرین ۱۲.۵ چگونه می‌توان تمرین ۱۱.۵ را تعمیم به حالتی داد که در آن قمارباز گیرنده یک کانال دلخواه Γ بوده، شرط‌بندی روی سمبل‌های ورودی صورت گرفته، و یک شرط موفق روی سمبل a_i، با احتمال p_i، منجر به بازیافت $۱/p_i$ برابر گرو می‌شود؟ چه می‌شد اگر قرار می‌گذاشتیم که یک شرط موفق روی سمبل a_i منجر به بازیافت $۱/p'_i$ برابر گرو می‌شد که در آن $\sum_i p'_i = ۱$ و $p'_i > ۰$ ؟ قمارباز با این روش برنده می‌بود یا بازنده؟

فصل ٦

کدهای تصحیح کننده خطا

صداقت شرط مالکیت خویش است.(Hamlet)

هدف ما اکنون ساخت کدهای C با نرخ ارسال خوب R و احتمال خطای پایین Pr_E است که وجود آنها به‌وسیله قضیه اساسی شانون ثابت شد (بخش ۴.۵). این بخش از موضوع تحت عنوان نظریه کدگذاری (یا کدهای تصحیح کننده خطا) قرار می‌گیرد، در برابر نظریه اطلاعات که در بر گیرنده مطالب فصول قبل می‌باشد. ساخت یک چنین کدهایی کار خیلی دشواری بوده، و ما روی چند مثال ساده تمرکز خواهیم کرد تا بعضی روش‌های به‌کار رفته در ساخت کدهای پیشرفته‌تر را توضیح دهیم.

۱.٦ مفاهیم مقدماتی

از حالا به بعد فرض بر این است که از یک کانال Γ استفاده می‌کنیم که همچون کانال BSC در آن الفبای ورودی و خروجی A و B یکسان هستند. این فرض هیچ محدودیتی بر کلیت مسئله ایجاد نمی‌کند زیرا اگر این خاصیت برقرار نباشد می‌توان A و B را با $A \cup B$ عوض کرد. این الفبای مشترک متناهی را با F نمایش می‌دهیم، زیرا اغلب از میدان استفاده خواهیم کرد تا بتوان روش‌های جبری را به‌کار برد. برای این‌که F میدان باشد باید تحت چهار عمل جمع، تفریق، ضرب و تقسیم (بر اعضای ناصفر) بسته بوده و اصول موضوعه‌ای همچون $ab = ba$ و $a(b + c) = ab + ac$ برقرارباشند. مثال‌های استاندارد حاوی میدان‌های $\mathbf{Q}$، $\mathbf{R}$ و $\mathbf{C}$ هستند که به‌ترتیب میدان اعداد گویا، حقیقی، و مختلط می‌باشند. این میدان‌ها نامتناهی هستند، و ما برای مقاصد خود نیاز به میدان‌های متناهی همچون $\mathbf{Z}_p$، میدان اعداد

صحیح به هنگ p، داریم که p عددی اول است. نتیجه اساسی که از میدان‌های متناهی نیاز داریم قضیه زیر است.

قضیه ۱.۶

(الف) یک میدان متناهی از مرتبه q وجود دارد اگر و فقط اگر $q = p^e$ که p اول بوده و e عددی طبیعی است.

(ب) هر دو میدان هم‌مرتبه ایزومورف هستند.

در بسیاری از کتاب‌های درسی جبری (مانند [KR 83]) اثبات این نتیجه موجود است و بنابراین آن‌را بدون اثبات می‌پذیریم. میدان یکتای از مرتبه q به نام میدان گالوای F_q یا $GF(q)$ معروف است. اگر $e = ۱$ آنگاه $q = p$ و بنابراین $F_q = F_p = \mathbf{Z}_p$، که میدان اعداد صحیح به هنگ عدد اول p می‌باشد. با این وجود اگر $e > ۱$ آنگاه q مرکب بوده و در این حالت $\mathbf{Z}_q$ میدان نیست: به‌عنوان نمونه p^e در $\mathbf{Z}_q$ برابر ۰ است اگرچه $p \neq ۰$، و بنابراین p یک مقسوم‌علیه ۰ است. این بدین معنی است که اگر $e > ۱$ آنگاه $F_q \neq \mathbf{Z}_q$. درعوض می‌توان F_q را میدان حاصل از الحاق یک ریشه α از چندجمله‌ای $f(x)$ به $\mathbf{Z}_p$ تعریف کرد که $f(x)$ یک چندجمله‌ای تحویل‌ناپذیر از درجه e روی $\mathbf{Z}_p$ است. این مشابه به‌دست آوردن میدان $\mathbf{C}$ از روی $\mathbf{R}$ با ضمیمه نمودن ریشه $i = \sqrt{-۱}$ از چندجمله‌ای $f(x) = ۱ + x^۲$ به $\mathbf{R}$ می‌باشد. در این‌صورت اعضای F_q به‌فرم $a_۰ + a_۱\alpha + \cdots + a_{e-۱}\alpha^{e-۱}$ هستند که $a_۰, a_۱, \ldots a_{e-۱} \in \mathbf{Z}_p$، و عمل جمع و ضرب به هنگ $f(\alpha)$ انجام می‌شوند.

مثال ۲.۶ چندجمله‌ای $f(x) = ۱ + x + x^۲$ در میدان $\mathbf{Z}_۲$ ریشه ندارد (چون $f(۰) = f(۱) = ۱$)، پس هیچ عامل خطی نداشته و بنابراین تحویل‌ناپذیر است. یک ریشه از $f(x)$ مانند α را به $\mathbf{Z}_۲$ ضمیمه می‌کنیم. در این‌صورت میدان چهار عضوی $F_۴$ حاصل می‌شود

$$F_۴ = \{a + b\alpha \mid a, b \in \mathbf{Z}_۲\} = \{۰, ۱, \alpha, ۱ + \alpha\}$$

که در آن $\alpha^۲ + \alpha + ۱ = ۰$ و بنابراین $\alpha^۲ = -۱ - \alpha = ۱ + \alpha$. برای نمونه، $\alpha(۱ + \alpha) = \alpha + \alpha^۲ = ۱ + ۲\alpha = ۱$ و از این‌رو α و $۱ + \alpha$ معکوس ضربی یکدیگر در $F_۴$ هستند. برای ساخت مشابه‌ای از میدان‌های متناهی تمرینات تکمیلی ۱۶.۶ و ۱۷.۶ را ببینید.

ساختار دقیق F_q معمولاً برای اهداف ما مهم نبوده و تنها کافی است که وجود آن‌را برای توان اعداد اول بپذیریم. با این وجود، کدهای پیشرفته‌تری وجود دارند که از حوصله این

کتاب خارج بوده و برای فهم آنها به دانش عمیق‌تری درباره میدان‌های متناهی نیاز است. محاسبات در F_q شبیه به آن در هر میدان دیگری است با این تفاوت که اگر $q = p^e$ آنگاه در F_q p برابر ۰ است. همچنین یک رابطه ترتیبی طبیعی $<$ مانند آنچه در $\mathbf{R}$ و $\mathbf{Q}$ مطرح است در F_q وجود ندارد. در بسیاری از حالات روی کدهای دوتایی تمرکز خواهیم کرد و از این‌رو $F_q = \mathbf{Z}_2 = \{0, 1\}$ با $1 + 1 = 0$.

از این‌جا به بعد از کدهای قالبی استفاده خواهیم کرد، کدهایی که در آنها تمامی کدکلمه‌ها هم‌طول هستند. این تناقضی با استفاده قبلی ما از کدهای با طول متغیر به‌منظور کارایی ندارد. می‌توان ابتدا از کدگذاری منبع استفاده کرده و سپس دنباله حاصل را به قالب‌های به‌طول k تقسیم نمود. این قالب‌های به‌طول k را با کدکلمه‌های به‌طول n نمایش خواهیم داد. سعی می‌شود که کدکلمه‌های به‌طول n از نقطه‌نظر فاصله همینگ تا حد ممکن از هم دور باشند به‌قسمی‌که کد حاصل از نظر تصحیح خطا خوب باشد.

اگر از کدکلمه‌های به‌طول n استفاده کنیم، آنگاه یک کد $\mathcal{C}$ به‌طول n یک زیرمجموعه از مجموعه $\mathcal{V} = F^n$ است. اگر F میدان باشد آنگاه $\mathcal{V}$ یک فضای برداری به بعد n روی F است که اعمال آن جمع و ضرب اسکالر نقطه‌به‌نقطه است؛ اگر $\mathbf{u} = u_1 \cdots u_n$ و $\mathbf{v} = v_1 \cdots v_n$ در $\mathcal{V}$ بوده و $a, b \in F$، آنگاه $a\mathbf{u} + b\mathbf{v}$ کلمه‌ای است که مختص i ام آن $au_i + bv_i$ می‌باشد. گوییم $\mathcal{C}$ یک کد خطی (یا یک کد گروهی) است اگر $\mathcal{C}$ یک زیر فضای خطی $\mathcal{V}$ باشد: یعنی این‌که $\mathcal{C}$ غیرخالی بوده و اگر $\mathbf{u}, \mathbf{v} \in \mathcal{C}$ آنگاه برای هر $a, b \in F$ داشته باشیم $a\mathbf{u} + b\mathbf{v} \in \mathcal{C}$. بالاخص هر کد خطی شامل بردار صفر $\mathbf{0} = 00 \ldots 0$ است زیرا برای هر $\mathbf{u}, \mathbf{v} \in \mathcal{C}$ داریم $0\mathbf{u} + 0\mathbf{v} = \mathbf{0}$.

تمرین ۱.۶ نشان دهید اگر $\mathcal{C}$ و $\mathcal{C}'$ دو کد خطی در $\mathcal{V}$ باشند، آنگاه کدهای $\mathcal{C} \cap \mathcal{C}'$ و $\mathcal{C} + \mathcal{C}' = \{\mathbf{u} + \mathbf{u}' | \mathbf{u} \in \mathcal{C}, \mathbf{u}' \in \mathcal{C}'\}$ نیز خطی هستند. تحت چه شرایطی $\mathcal{C} \cup \mathcal{C}'$ یک کد خطی است؟

اکثر کدها غیرخطی هستند از این جهت که درصد پایینی از زیرمجموعه‌های $\mathcal{V}$ زیرفضا هستند؛ با این وجود اکثر کدهایی که مطالعه و به‌کار می‌روند خطی هستند زیرا از فهم و به‌کارگیری ساده‌تری برخوردار می‌باشند. می‌توان مشابه قضیه اساسی شانون را برای کدهای خطی ثابت کرد: به‌عوض انتخاب یک کد تصادفی $\mathcal{C} \subseteq \mathcal{V}$، یک زیرمجموعه تصادفی از $\mathcal{V}$ را انتخاب و آن‌را به‌عنوان مبنای یک کد خطی $\mathcal{C} \subseteq \mathcal{V}$ درنظر گرفته و نشان می‌دهیم که $\mathcal{C}$ با افزایش n واجد خواص لازم می‌باشد.

همیشه $|\mathcal{C}|$ را با M نمایش خواهیم داد. وقتی $\mathcal{C}$ خطی است داریم $M = q^k$ که $k = \dim(\mathcal{C})$ بعد زیرفضای $\mathcal{C}$ می‌باشد؛ این بدین جهت است که هر عضو $\mathcal{C}$ یک نمایش منحصربه‌فرد $a_1\mathbf{u}_1 + \cdots + a_k\mathbf{u}_k$ دارد که در آن $a_i \in F$ و $\mathbf{u}_1, \ldots, \mathbf{u}_k$ یک مبنای $\mathcal{C}$ بوده،

و برای هر یک از ضرایب تعداد $q = |F|$ انتخاب وجود دارد. در این‌صورت $\mathcal{C}$ را یک $[n, k]$–کد خطی می‌نامیم.

نرخ یک کد $\mathcal{C}$ برابر است با

$$R = \frac{\log_q M}{n}, \tag{۱.۶}$$

بنابراین برای یک $[n, k]$–کد خطی داریم

$$R = \frac{k}{n}. \tag{۲.۶}$$

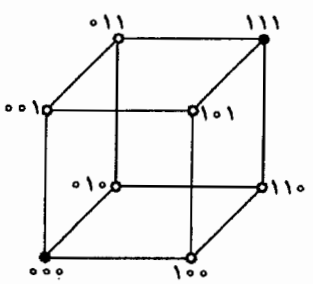
می‌توان این‌را چنین تعبیر کرد که از هر کدکلمه به‌طول n تعداد k بیت به‌عنوان بیت‌های اطلاعات مطرح هستند که در واقع حامل اطلاعاتی هستند که می‌خواهیم ارسال کنیم، و بقیه $n - k$ بیت به‌عنوان بیت‌های کنترل می‌باشند که نقش تاییدی یا حفاظتی برای بخش اطلاعات دارند.

فرض می‌کنیم که تمامی کدکلمه‌های $\mathcal{C}$ دارای احتمال یکسان بوده و برای کدگشایی از قاعده نزدیک‌ترین همسایه (نسبت به فاصله همینگ) استفاده می‌شود.

۲.۶ نمونه‌هایی از یک کد

در این‌جا به‌عنوان مثال به بررسی چند کد می‌پردازیم. فهم آنها ساده است ولی از نظر نرخ یا احتمال خطا چندان مؤثر نیستند؛ مثال‌های مؤثرتری را در بخش‌های بعد ملاحظه خواهیم کرد.

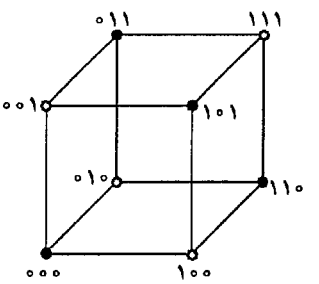
مثال ۳.۶ کد تکراری $\mathcal{R}_n$ روی F متشکل از کلمات $\mathbf{u} = uu \dots u \in \mathcal{V} = F^n$ است که $u \in F$، $M = |F| = q$. اگر F یک میدان باشد آنگاه $\mathcal{R}_n$ یک کد خطی به بعد $k = ۱$ است که توسط کلمه (یا بردار) ۱ ... ۱۱ تولید می‌شود. شکل ۱.۶ کد دوتایی $\mathcal{R}_۳$ را به‌عنوان یک زیرمجموعه از $\mathcal{V} = F_۲^۳$ نشان می‌دهد که در آن کدکلمه‌ها با رئوس توپر مشخص شده‌اند.



شکل ۱.۶

در فصل ۵ دیدیم که وقتی $q = ۲$ و n فرد است، کد $\mathcal{R}_n$ تعداد $(n-۱)/۲$ خطا را تصحیح می‌کند؛ منظور ما این است که اگر یک کدکلمه $\mathbf{u} \in \mathcal{R}_n$ ارسال شود و حداکثر $(n-۱)/۲$ از n سمبل نادرست ارسال شوند آنگاه کدگشایی نزدیک‌ترین همسایه همیشه درست عمل می‌کند. بحث مشابه‌ای به‌سادگی نشان می‌دهد که برای هر q و n اگر کدگشایی نزدیک‌ترین همسایه را به‌کار ببریم آنگاه $\mathcal{R}_n$ تعداد $\lfloor (n-۱)/۲ \rfloor$ خطا را تصحیح می‌کند که $\lfloor x \rfloor = \max\{m \in \mathbf{Z} \mid m \leq x\}$. این عالی است، ولی متأسفانه (۲.۶) ایجاب می‌کند که نرخ $\mathcal{R}_n$ برابر $R = ۱/n$ بوده و اگر $n \longrightarrow \infty$ آنگاه $R \longrightarrow ۰$، که این بد است.

مثال ۴.۶ کد کنترل تساوی[۱] $\mathcal{P}_n$ روی میدان F_q متشکل است از تمامی بردارهای $\mathbf{u} = u_۱u_۲\ldots u_n \in \mathcal{V}$ به‌قسمی‌که $\sum_i u_i = ۰$؛ می‌توان مؤلفه‌های $u_۱, \ldots, u_{n-۱}$ را به‌عنوان ارقام اطلاعات و u_n را به‌عنوان رقم کنترل درنظر گرفت. به‌عنوان نمونه، اگر $n = ۳$ و $q = ۲$ آنگاه $\mathcal{P}_۳ = \{۰۰۰, ۰۱۱, ۱۰۱, ۱۱۰\}$. شکل ۲.۶ این کد را نمایش می‌دهد.



شکل ۲.۶

چون این کد با معادله خطی $\sum_i u_i = ۰$ تعریف می‌شود یک کد خطی است. بعد آن $k = n - ۱$ بوده و بردارهای $\mathbf{u}_{n-۱} = \mathbf{e}_{n-۱} - \mathbf{e}_n, \ldots, \mathbf{u}_۱ = \mathbf{e}_۱ - \mathbf{e}_n$ تشکیل یک مبنا برای این کد می‌دهند که $\mathbf{e}_i$ ها بردارهای مبنای استاندارد هستند. (برای دیدن این‌که این بردارها یک مبنای $\mathcal{P}_n$ هستند ملاحظه کنید که هر بردار $\mathbf{u} = u_۱\ldots u_n \in \mathcal{P}_n$ به‌صورت منحصربه‌فردی به‌شکل ترکیب خطی $u_۱\mathbf{u}_۱ + \cdots + u_{n-۱}\mathbf{u}_{n-۱}$ نوشته می‌شود.) پس $M = q^{n-۱}$ و $R = (n-۱)/n$. بنابراین اگر $n \longrightarrow \infty$ آنگاه $R \longrightarrow ۱$، که نشان می‌دهد از نقطه‌نظر نرخ ارسال کد $\mathcal{P}_n$ خوب است.

متأسفانه از نقطه‌نظر تصحیح خطا این کد تقریباً بی‌فایده است، زیرا می‌تواند تنها یک خطا را کشف کند بدون این‌که توانایی اصلاح آن‌را داشته باشد. فرض کنید $\mathbf{u} = u_۱\ldots u_n \in \mathcal{P}_n$ ارسال شده و $\mathbf{v} = v_۱\ldots v_n \in \mathcal{V}$ دریافت شود. گیرنده $\sum_i v_i$ را در F_q

[۱] parity check code

محاسبه می‌کند. اگر یک خطا رخ داده باشد آنگاه دقیقاً یک رقم v_i در $\mathbf{v}$ با رقم متناظرش در $\mathbf{u}$ تفاوت دارد؛ چون $\sum_i u_i = 0$ پس $\sum_i v_i \neq 0$، و بنابراین گیرنده می‌داند که $\mathbf{v}$ یک کدکلمه نیست و باید خطا رخ داده باشد؛ ولی هیچ راهی برای پی‌بردن به این‌که کدام مؤلفه نادرست است وجود ندارد زیرا با تغییر هریک از مؤلفه‌ها می‌توان به یک کدکلمه دست یافت. بدتر از این وجود تعداد زوج خطا است که قابل کشف نیست.

مثال ۵.۶ کد دوتایی همینگ $\mathcal{H}_7$ یک کد خطی به‌طول $n = 7$ روی F_2 است. این از جمله اولین کدهای تصحیح کننده خطا است که در سال ۱۹۴۷ توسط همینگ معرفی شد و سپس در آزمایشگاه Bell گسترش داده شد [Sh48،Ha50،Ha48]. (برای یک گزارش جذاب راجع به روزهای اولیه نظریه کدگذاری به [Th 83] مراجعه کنید).

برای ساختن این کد از شکل ۳.۶ استفاده می‌کنیم که یک دیاگرام ون را برای سه مجموعه A، B و C نشان می‌دهد. نواحی متناظر با مجموعه‌های $A\cap B\cap \overline{C}$، $A\cap \overline{B}\cap C$، $\overline{A}\cap B\cap C$، $A\cap \overline{B}\cap \overline{C}$، $\overline{A}\cap B\cap \overline{C}$، $\overline{A}\cap \overline{B}\cap C$، $A\cap B\cap C$ به‌ترتیب با اعداد ۱، ۲، ...، ۷ شماره‌گذاری می‌شوند (مجموعه $\overline{A}\cap \overline{B}\cap \overline{C}$ را درنظر نمی‌گیریم)؛ بنابراین i در نواحی A، B یا C قرار می‌گیرد بسته به این‌که در نمایش دوتایی abc از عدد i اعداد a، b و c برابر ۱ باشند. برای نمونه، عدد ۵ به‌صورت ۱۰۱ در مبنای ۲ نوشته می‌شود و از این‌رو ۵ متناظر با مجموعه $A\cap \overline{B}\cap C$ بوده در حالی‌که $\overline{A}\cap B\cap \overline{C}$ متناظر با ۰۱۰ یعنی ۲ می‌باشد.

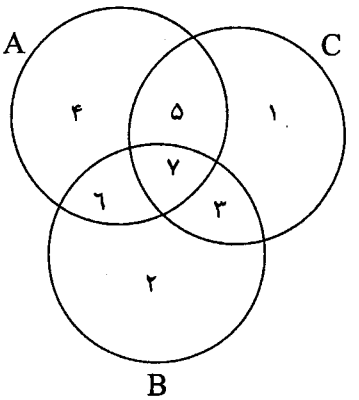
برای کدکردن یک دنباله دوتایی به‌طول چهار $\mathbf{a} = a_1a_2a_3a_4$ از دنباله اطلاعات به کدکلمه $\mathbf{u} = u_1 \dots u_7$، ابتدا قرار می‌دهیم $u_3 = a_1$، $u_5 = a_2$، $u_6 = a_3$، $u_7 = a_4$؛ این اعداد را به‌ترتیب در نواحی با شماره ۳، ۵، ۶، ۷ قرار می‌دهیم (شکل ۳.۶). مقدار u_4 را برابر ۰ یا ۱ تعریف کرده و آن‌را در ناحیه ۴ می‌نویسیم به‌قسمی‌که مجموع ۴ عدد قرار گرفته در مجموعه A در میدان F_2 صفر باشد:

$$u_4 + u_5 + u_6 + u_7 = 0.$$

مؤلفه‌های u_1 و u_2 را مشابهاً با به‌کار بردن مجموعه‌های C و B تعریف می‌کنیم، به‌طوری‌که

$$u_2 + u_3 + u_6 + u_7 = 0,$$
$$u_1 + u_3 + u_5 + u_7 = 0.$$

(توجه کنید اندیس‌های ظاهر شده در این سه معادله به‌ترتیب آنهایی هستند که نمایش آنها در مبنای ۲ حاوی ۱ در موقعیت‌های اول، دوم و سوم، یعنی a، b، و c است.)



شکل ۳.۶

کد $\mathcal{H}_۷$ متشکل از تمامی کلمات $\mathbf{u} \in \mathcal{V} = F_۲^۷$ است که به این طریق ساخته می‌شوند. چون $\mathcal{H}_۷$ براساس معادلات خطی بین متغیرهای u_i تعریف شده است یک کد خطی است. تعداد $۲^۴ = ۱۶$ انتخاب برای $a_۱a_۲a_۳a_۴$ وجود داشته و هر یک از این قالب‌ها به‌صورت منحصربه‌فردی کدکلمه $\mathbf{u} = u_i \dots u_۷$ را تعیین می‌کند پس $M = |\mathcal{H}_۷| = ۱۶$؛ این همچنین نشان می‌دهد بعد $\mathcal{H}_۷$ برابر $k = ۴$ بوده و بردارهای $\mathbf{u}_۱ = ۱۱۱۰۰۰۰$، $\mathbf{u}_۲ = ۱۰۰۱۱۰۰$، $\mathbf{u}_۳ = ۰۱۰۱۰۱۰$، $\mathbf{u}_۴ = ۱۱۰۱۰۰۱$ تشکیل یک مبنا برای $\mathcal{H}_۷$ می‌دهند. این چهار بردار به‌ترتیب با انتخاب $a_i = ۱$، $۴ \le i \le ۱$، و $a_{j \neq i} = ۰$ ایجاد می‌شوند.

این کد توان این‌را دارد که هر خطای به وزن یک (یک خطا) را تصحیح کند. فرض کنید $\mathbf{u} \in \mathcal{H}_۷$ ارسال و $\mathbf{v} \in \mathcal{V}$ دریافت شده است به‌قسمی‌که تفاوت $\mathbf{u}$ و $\mathbf{v}$ دقیقاً در i امین رقم می‌باشد. گیرنده مقادیر زیر را در $F_۲$ محاسبه می‌کند

$$s_۱ = v_۴ + v_۵ + v_۶ + v_۷,$$
$$s_۲ = v_۲ + v_۳ + v_۶ + v_۷,$$
$$s_۳ = v_۱ + v_۳ + v_۵ + v_۷.$$

اینها براساس تعریف $\mathcal{H}_۷$ همگی باید ۰ باشند، ولی وجود خطا باعث می‌شود که حداقل یکی از آنها ۱ باشد. اکنون v_i در معادله مربوط به s_j، $۳ \le j \le ۱$، ظاهر می‌شود اگر و فقط اگر j امین رقم بسط i در مبنای ۲ برابر ۱ باشد. از این‌رو مقدار s_j برابر j امین رقم i است. بنابراین $\mathbf{s} = s_۱s_۲s_۳$ نمایش دوتایی i است. پس گیرنده با محاسبه $s_۱$، $s_۲$، و $s_۳$ محل رقم نادرست v_i را معین نموده و با تعریف $u_i = v_i + ۱$ آن‌را تصحیح می‌کند.

برای نمونه، فرض کنید می‌خواهیم $\mathbf{a} = ۰۱۱۰$ را کدگذاری کنیم. رقم‌های اطلاعات را تعریف می‌کنیم $u_۳ = ۰$، $u_۵ = u_۶ = ۱$، $u_۷ = ۰$؛ سپس سه معادله خطی را حل کرده و رقم‌های کنترل را تعیین می‌کنیم: $u_۴ = ۰$، $u_۲ = ۱$، $u_۱ = ۱$؛ بنابراین کدکلمه ارسال شده برابر است با $\mathbf{u} = ۱۱۰۰۱۱۰ = \mathbf{u}_۲ + \mathbf{u}_۳$. حال فرض کنید در ارسال یک خطا در رقم سوم ایجاد شده و کلمه دریافتی $\mathbf{v} = ۱۱۱۰۱۱۰$ باشد. گیرنده مقادیر $s_۱$، $s_۲$، و $s_۳$ را محاسبه می‌کند: $s_۳ = ۰ + ۱ + ۰ = ۰$، $s_۲ = ۱ + ۱ + ۰ + ۱ = ۱$، $s_۱ = ۰ + ۱ + ۱ = ۰$؛ پس $\mathbf{s} = s_۱s_۲s_۳ = ۰۱۱$، که نمایش دوتایی $i = ۳$ است. پس رقم سوم $\mathbf{v}$ تغییر یافته و $\Delta(\mathbf{v}) = ۱۱۰۰۱۱۰$ به‌دست می‌آید که در واقع کلمه ارسال شده است. از این‌جا ارقام اطلاعات از روی موقعیت‌های ۳، ۵، ۶، ۷ خوانده می‌شوند: $\mathbf{a} = ۰۱۱۰$.

این کد هر یک خطایی را تصحیح می‌کند ولی اگر بیش از یک خطا رخ دهد خطای کدگشایی اتفاق می‌افتد. مثلاً، در مورد بالا فرض کنید در رقم‌های $u_۳$ و $u_۴$ خطا رخ دهد به‌قسمی‌که $\mathbf{v}' = ۱۱۱۱۱۱۰$ در یافت شود. گیرنده نتیجه می‌گیرد $s'_۱ = ۱$، $s'_۲ = ۱$، $s'_۳ = ۱$ و بنابراین $\mathbf{s}' = s'_۱s'_۲s'_۳ = ۱۱۱$، که معرف وجود خطا در موقعیت $i = ۷$ است، پس $\mathbf{v}'$ به $\Delta(\mathbf{v}') = ۱۱۱۱۱۱۱$ دکد می‌شود که درست نیست.

تمرین ۲.۶ کدکلمه‌ای از $\mathcal{H}_۷$ را پیدا کنید که معرف ارقام اطلاعاتی ۱۱۰۱ باشد، و نشان دهید چگونه یک خطا در ۶ امین سمبل آن تصحیح می‌شود. اگر در سمبل‌های چهارم و ششم خطا رخ دهد چه می‌شود؟

اگرچه کدهای دوتایی $\mathcal{R}_۳$ و $\mathcal{H}_۷$ هر دو یک خطا را تصحیح می‌کنند ولی نرخ $R = ۴/۷$ برای $\mathcal{H}_۷$ بسیار بهتر از نرخ $\frac{۱}{۳}$ برای $\mathcal{R}_۳$ است. در فصل ۷ ساخت $\mathcal{H}_۷$ را تعمیم داده و دنباله‌ای از کدهای دوتایی $\mathcal{H}_n$، $n = ۲^c - ۱$، به‌دست می‌آوریم که همگی یک خطا را تصحیح کرده و با $n \longrightarrow \infty$ نرخ $\mathcal{H}_n$ به ۱ میل می‌کند. ممکن است علاقمند باشید که پیشاپیش در مورد چگونگی ساخت این کدها با تعویض مجموعه‌های A، B و C با $A_۱$،...، A_c فکر کنید.

مثال ۶.۶ فرض کنید $\mathcal{C}$ یک کد به‌طول n روی یک میدان F باشد. در این‌صورت می‌توانیم یک کد به‌طول $n + ۱$ روی F بسازیم که کد بسط یافته $\overline{\mathcal{C}}$ نامیده می‌شود. برای ساخت $\overline{\mathcal{C}}$، به انتهای کدکلمه مفروض $\mathbf{u} = u_۱ \dots u_n \in \mathcal{C}$ رقم $u_{n+۱}$ را به‌قسمی اضافه می‌کنیم که $u_{n+۱} = -\sum_{i=۱}^{n} u_i$. واضح است که $|\overline{\mathcal{C}}| = |\mathcal{C}|$ و اگر $\mathcal{C}$ خطی باشد آنگاه $\overline{\mathcal{C}}$ نیز خطی بوده و بعد آن برابر بعد $\mathcal{C}$ است. مثلاً اگر $\mathcal{C} = \mathcal{V} = F^n$ آنگاه $\overline{\mathcal{C}} = \mathcal{P}_{n+۱} \subset F^{n+۱}$.

مثال ۷.۶ اگر $\mathcal{C}$ یک کد به‌طول n باشد یک کد پنچر شده $\mathcal{C}^\circ$ به‌طول $n - ۱$ از روی $\mathcal{C}$ به این ترتیب می‌سازیم که ابتدا یک مختص i انتخاب کرده و سپس سمبل i ام

تمامی کدکلمه‌های $\mathcal{C}$ را حذف می‌کنیم. در حالت کلی ساخت $\mathcal{C}^{\circ}$ به انتخاب مختص i بستگی دارد.

۳.۶ فاصله می‌نیمم

با به‌کار بردن کدگشایی نزدیک‌ترین همسایه می‌توان Pr_E را پایین نگهداشت اگر از کدکلمه‌هایی استفاده کنیم که از هم دور باشند، زیرا در این‌صورت با احتمال زیاد کلمه ارسال شده $\mathbf{u}$ نزدیک‌ترین کدکلمه $\Delta(\mathbf{v})$ به کلمه دریافت شده $\mathbf{v} \in \mathcal{V}$ است. از این‌رو می‌نیمم فاصله یک کد را چنین تعریف می‌کنیم

$$d = d(\mathcal{C}) = \min\{d(\mathbf{u}, \mathbf{u}') \mid \mathbf{u}, \mathbf{u}' \in \mathcal{C}, \mathbf{u} \neq \mathbf{u}'\}. \qquad (۳.۶)$$

یک کد به‌طول n و با M کدکلمه، و می‌نیمم فاصله d را معمولاً یک (n, M, d)–کد نامیده، و اگر کد خطی بوده و بعد آن k باشد آن‌را یک $[n, k, d]$–کد می‌نامیم.

هدف ما انتخاب کدهای $\mathcal{C}$ با می‌نیمم فاصله نسبتاً بزرگ d می‌باشد به‌قسمی‌که Pr_E کوچک باشد. اگر $\mathcal{C}$ دارای M کدکلمه باشد آنگاه پیدا کردن d به‌وسیله (۳.۶) مستلزم پیدا کردن و مقایسه $\binom{M}{۲} = M(M-۱)/۲$ فاصله است که می‌تواند کسل کننده باشد. با این حال، اگر $\mathcal{C}$ خطی باشد این کار خیلی ساده‌تر خواهد بود.

ابتدا وزن هر بردار $\mathbf{v} = v_۱ \dots v_n \in \mathcal{V}$ را برابر فاصله آن با بردار $\mathbf{0} = ۰۰\dots۰$ تعریف می‌کنیم

$$wt(\mathbf{v}) = d(\mathbf{v}, \mathbf{0}). \qquad (۴.۶)$$

به‌عبارت دیگر، وزن $\mathbf{v}$، $wt(\mathbf{v})$، برابر تعداد مؤلفه‌های ناصفر $\mathbf{v}$ تعریف می‌شود. به‌سادگی می‌توان دید که برای هر $\mathbf{u}, \mathbf{u}' \in \mathcal{V}$ رابطه $d(\mathbf{u}, \mathbf{u}') = wt(\mathbf{u} - \mathbf{u}')$ برقرار است.

لم ۸.۶ اگر $\mathcal{C}$ یک کد خطی باشد آنگاه کمترین فاصله آن برابر است با

$$d = \min\{wt(\mathbf{v}) \mid \mathbf{v} \in \mathcal{C}, \mathbf{v} \neq \mathbf{0}\}.$$

اثبات داریم $d(\mathbf{u}, \mathbf{u}') = wt(\mathbf{v})$ که $\mathbf{v} = \mathbf{u} - \mathbf{u}'$. چون $\mathcal{C}$ یک زیرفضای $\mathcal{V}$ است با تغییر $\mathbf{u}$ و $\mathbf{u}'$ روی زوج‌های متمایز در $\mathcal{C}$، تفاضل آنها $\mathbf{v} = \mathbf{u} - \mathbf{u}'$ روی تمامی اعضای ناصفر $\mathcal{C}$ تغییر می‌کند. نتیجه می‌شود $d(\mathcal{C})$، که می‌نیمم فاصله بین این چنین زوج‌هایی است، برابر با می‌نیمم این وزن‌های ناصفر است. ■

قوت این نتیجه در این است که براساس آن در کدهای خطی برای محاسبه می‌نیمم فاصله نیاز به محاسبه و مقایسه $M-۱$ عدد است درحالی‌که این عدد در کدهای غیرخطی

برابر $M(M-۱)/۲$ است. در بخش ۳.۷ خواهیم دید که برای محاسبه می‌نیمم فاصله کدهای خطی حتی روش‌های بهتری نیز وجود دارد.

تمرین ۳.۶ تمام کدکلمه‌های کد همینگ $\mathcal{H}_۷$ (مثال ۵.۶) را نوشته و با به‌کار بردن لم ۸.۶ نشان دهید می‌نیمم فاصله برابر ۳ است.

تمرین ۴.۶ نشان دهید اگر $\mathcal{C}$ یک کد دوتایی خطی با می‌نیمم فاصله d باشد، آنگاه بسته به زوج یا فرد بودن d، کد $\overline{\mathcal{C}}$ می‌نیمم فاصله d یا $d+۱$ دارد. کلمات کد بسط یافته $\overline{\mathcal{H}}_۷$ را نوشته و می‌نیمم فاصله آن‌را تعیین کنید.

حال چگونگی تأثیر می‌نیمم فاصله یک کد را روی قابلیت تصحیح خطای آن بررسی می‌کنیم. گوییم کد $\mathcal{C}$ تعداد t خطا را تصحیح می‌کند، یا t خطا تصحیح کننده است، اگر هر وقت که یک کدکلمه $\mathbf{u} \in \mathcal{C}$ ارسال شود و حداکثر t خطا روی آن صورت بگیرد، آنگاه کلمه دریافتی $\mathbf{v}$ به $\mathbf{u}$ کدگشایی شود؛ معادلاً، اگر کلمات $\mathbf{u} \in \mathcal{C}$ و $\mathbf{v} \in \mathcal{V}$ در شرط $d(\mathbf{u},\mathbf{v}) \leq t$ صدق کنند، آنگاه از قاعده تصمیم Δ نتیجه شود $\Delta(\mathbf{v}) = \mathbf{u}$.

مثال ۹.۶ یک کد تکراری $\mathcal{R}_۳$ (روی هر الفبایی) یک خطا را تصحیح می‌کند ولی توان تصحیح بیش از یکی را ندارد (بخش ۲.۵ را برای حالت $q = ۲$ ملاحظه کنید). به‌عنوان نمونه، اگر $\mathbf{u} = ۱۱۱$ ارسال شده و $\mathbf{v} = ۱۰۱$ دریافت شود، آنگاه با قاعده نزدیک‌ترین همسایه داریم $\Delta(\mathbf{v}) = ۱۱۱ = \mathbf{u}$. ولی اگر $\mathbf{v} = ۰۰۱$ دریافت شود (یعنی دو خطا وجود دارد)، آنگاه $\Delta(\mathbf{v}) = ۰۰۰ \neq \mathbf{u}$.

اگر $\mathbf{u}$ ارسال و $\mathbf{v}$ دریافت شود آنگاه $\mathbf{e} = \mathbf{v} - \mathbf{u}$ را قالب خطا (یا بردار خطا) می‌نامیم زیرا درایه‌های ناصفر آن موقعیت خطاهای صورت گرفته در هنگام ارسال و مقدار این خطاها را مشخص می‌کنند. معادله $\mathbf{v} = \mathbf{u} + \mathbf{e}$ نشان می‌دهد که $\mathbf{v}$ متشکل از کلمه ارسالی و بردار خطا می‌باشد. تعداد سمبل‌های نادرست برابر $d(\mathbf{u},\mathbf{v}) = wt(\mathbf{e})$ بوده، و از این‌رو یک کد مفروض تعداد t خطا را اصلاح می‌کند اگر و فقط اگر بتواند تمامی بردارهای خطای $\mathbf{e} \in \mathcal{V}$ به وزن $wt(\mathbf{e}) \leq t$ را تصحیح کند.

قضیه ۱۰.۶ یک کد $\mathcal{C}$ با می‌نیمم فاصله d تعداد t خطا را تصحیح می‌کند اگر و فقط اگر $d \geq ۲t+۱$. (معادلاً، $\mathcal{C}$ تا $\lfloor (d-۱)/۲ \rfloor$ خطا را تصحیح می‌کند.)

اثبات $\Longleftarrow$ فرض کنید $\mathcal{C}$ دارای می‌نیمم فاصله $d \geq ۲t+۱$ باشد. فرض کنید $\mathbf{u}$ ارسال و $\mathbf{v} = \mathbf{u} + \mathbf{e}$ دریافت شده است که بردار خطای $\mathbf{e}$ دارای وزن $wt(\mathbf{e}) \leq t$ است، پس $d(\mathbf{u},\mathbf{v}) \leq t$. برای تمامی $\mathbf{u}' \neq \mathbf{u}$ در $\mathcal{C}$ داریم

$$d(\mathbf{u},\mathbf{u}') \geq d \geq ۲t+۱.$$

حال از نامساوی مثلث نتیجه می‌شود

$$d(\mathbf{u},\mathbf{u}') \le d(\mathbf{u},\mathbf{v}) + d(\mathbf{v},\mathbf{u}'),$$

پس

$$d(\mathbf{v},\mathbf{u}') \ge d(\mathbf{u},\mathbf{u}') - d(\mathbf{u},\mathbf{v}) \ge (2t+1) - t = t+1 > d(\mathbf{u},\mathbf{v}).$$

از این‌رو $\Delta(\mathbf{v}) = \mathbf{u}$، پس عمل کدگشایی درست است، و $\mathcal{C}$ تعداد t خطا را تصحیح می‌کند.
($\Longrightarrow$) فرض کنید $\mathcal{C}$ دارای می‌نیمم فاصله $d < 2t+1$ باشد، پس $d \le 2t$. می‌توان کلمات $\mathbf{u},\mathbf{u}' \in \mathcal{C}$ را به‌قسمی انتخاب کرد که $d(\mathbf{u},\mathbf{u}') = d$. در این‌صورت برداری چون $\mathbf{v} \in \mathcal{V}$ موجود است که

$$d(\mathbf{u},\mathbf{v}) \le t, \qquad d(\mathbf{u}',\mathbf{v}) \le t.$$

(به‌عنوان نمونه، در سمبل‌هایی که $\mathbf{u}$ و $\mathbf{u}'$ متفاوت هستند تعداد $\lfloor d/2 \rfloor$ از سمبل‌های $\mathbf{u}$ را تبدیل به سمبل‌های متناظر در $\mathbf{u}'$ می‌کنیم، کلمه حاصل $\mathbf{v}$ واجد خاصیت فوق است.) اکنون $\Delta(\mathbf{v})$ نمی‌تواند هر دو کدکلمه $\mathbf{u}$ و $\mathbf{u}'$ باشد، بنابراین حداقل یکی از این دو کلمه در صورت ارسال و دریافت به‌عنوان $\mathbf{v}$ به‌صورت نادرست دکد می‌شود. پس $\mathcal{C}$ تعداد t خطا را تصحیح نمی‌کند. ■

مثال ۱۱.۶ یک کد تکراری $\mathcal{R}_n$ به‌طول n دارای می‌نیمم فاصله $d = n$ است، زیرا برای هر دو کدکلمه متمایز رابطه $d(\mathbf{u},\mathbf{u}') = n$ برقرار است. بنابراین این کد $t = \lfloor (n-1)/2 \rfloor$ خطا را تصحیح می‌کند.

مثال ۱۲.۶ تمرین ۳.۶ نشان می‌دهد که کد همینگ $\mathcal{H}_7$ دارای می‌نیمم فاصله $d = 3$ است، بنابراین $t = 1$. مشابهاً، $\overline{\mathcal{H}}_7$ دارای می‌نیمم فاصله $d = 4$ (تمرین ۴.۶) است؛ از این‌رو این کد نیز دارای $t = 1$ است.

مثال ۱۳.۶ یک کد کنترل تساوی $\mathcal{P}_n$ به‌طول n دارای می‌نیمم فاصله $d = 2$ است. به‌عنوان مثال فاصله دو کلمه $\mathbf{u} = 1(-1)0\ldots0$ و $\mathbf{u}' = 0 = 00\ldots0$ برابر ۲ است، ولی هیچ دو کلمه‌ایی به فاصله ۱ از یکدیگر نیستند. نتیجه می‌شود که تعداد خطاهای قابل تصحیح به‌وسیله $\mathcal{P}_n$ برابر $t = \lfloor (n-1)/2 \rfloor = 0$ است: برای مثال کلمه $\mathbf{v} = 10\ldots0$ می‌تواند به‌عنوان $\mathbf{u}$ یا $\mathbf{u}'$ دکد شود که هر یک از آنها با یک خطا منجر به $\mathbf{v}$ می‌شوند.

اگرچه $\mathcal{P}_n$ کاربردی در تصحیح خطا ندارد ولی یک خطا را کشف می‌کند. در حالت کلی، فرض کنید که $\mathcal{C}$ دارای می‌نیمم فاصله d باشد، نیز فرض کنید یک کدکلمه $\mathbf{u}$ ارسال

شده و کلمه $\mathbf{v} = \mathbf{u} + \mathbf{e}$ دریافت شده است که $1 \leq wt(\mathbf{e}) \leq d - 1$؛ در این‌صورت چون $0 < d(\mathbf{u}, \mathbf{v}) < d$ کلمه $\mathbf{v}$ نمی‌تواند یک کدکلمه باشد، از این‌رو گیرنده می‌داند که در کلمه دریافتی $\mathbf{v}$ حداقل یک خطا وجود دارد. با این حال اگر $wt(\mathbf{e}) = d$ آنگاه کلمه دریافتی $\mathbf{v}$ ممکن است یک کدکلمه باشد که در این حالت گیرنده نمی‌داند که آیا $\mathbf{v}$ یک کدکلمه درست ارسال شده است و یا کلمه نادرست ارسال شده. بنابراین می‌گوییم $\mathcal{C}$ تعداد $d - 1$ خطا را کشف می‌کند.

مثال ۱۴.۶ کدهای $\mathcal{R}_n$ و $\mathcal{P}_n$ به‌ترتیب دارای می‌نیمم فاصله n و ۲ هستند، پس $\mathcal{R}_n$ تعداد $n - 1$ خطا را کشف می‌کند حال آن‌که $\mathcal{P}_n$ فقط یک خطا را؛ $\mathcal{H}_7$ دارای $d = 3$ است و از این‌رو تعداد ۲ خطا را کشف می‌کند.

۴.۶ کران کره‌چینی

ملاحظه کرده‌ایم که یک کد $\mathcal{C}$ با می‌نیمم فاصله d تعداد $t = \lfloor (d-1)/2 \rfloor$ خطا را تصحیح می‌کند. کره‌های

$$S_t(\mathbf{u}) = \{\mathbf{v} \in \mathcal{V} \mid d(\mathbf{u}, \mathbf{v}) \leq t\} \quad (\mathbf{u} \in \mathcal{C}) \tag{۵.۶}$$

دوبه‌دو متمایز بوده و هر $S_t(\mathbf{u})$ متشکل است از تمامی بردارهای $\mathbf{v}$ که به $\mathbf{u}$ دکد می‌شوند. برای به‌دست آوردن کدهای خوب تصحیح کننده خطا این شعاع مشترک t باید بزرگ باشد. ولی برای داشتن نرخ خوب در ارسال داده‌ها یعنی $R = (\log_q M)/n$، نیاز به این است که M بزرگ باشد که M تعداد این کره‌ها است. اگر n و q ثابت باشند آنگاه چون کره‌ها جدا از هم هستند این دو هدف در تعارض با هم قرار دارند. می‌توان $\mathcal{V}$ را به‌عنوان یک جعبه n بعدی با ابعاد $q \times q \times \cdots \times q$ درنظر گرفت که می‌خواهیم تعداد زیادی از کره‌های جدا از هم بزرگ را در آن جا دهیم. واضح است که در میزان رسیدن به این هدف محدودیتی وجود دارد. نتیجه بعدی که به کران کره‌چینی همینگ[۱] [Ha 50] معروف است این حد را با دقت بیان می‌کند.

قضیه ۱۵.۶ فرض کنید $\mathcal{C}$ یک کد q تایی t خطا تصحیح کننده به‌طول n و با M کدکلمه باشد. در این‌صورت

$$M\left(1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t\right) \leq q^n.$$

اثبات با احتساب یک کره برای هر کدکلمه $\mathbf{u} \in \mathcal{C}$ تعداد M کره $S_t(\mathbf{u}) \subseteq \mathcal{V}$ وجود دارد.

[۱] Haming's sphere - packing bound

همچون تمرین ۴.۵، برای هر $\mathbf{u} \in C$ و هر i، تعداد بردارهای $\mathbf{v} \in \mathcal{V}$ با خاصیت $d(\mathbf{u}, \mathbf{v}) = i$ برابر است با $\binom{n}{i}(q-1)^i$: یک بردار $\mathbf{v}$ این چنینی باید در دقیقاً i مختص خود با $\mathbf{u}$ تفاوت داشته باشد؛ این مؤلفه‌ها به $\binom{n}{i}$ طریق انتخاب شده، و برای هر انتخاب در هر مختص $q-1$ انتخاب برای مقدار بردار مطلوب $\mathbf{v}$ وجود دارد. با جمع نمودن این اعداد برای $1 \leq i \leq t$ و ملاحظه (۵.۶) نتیجه می‌شود برای هر $\mathbf{u} \in C$ رابطه زیر برقرار است.

$$|S_t(\mathbf{u})| = 1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t. \qquad (6.6)$$

حال این M کره با توجه به $2t < d$ جدا از هم بوده و همگی در $\mathcal{V}$، $|\mathcal{V}| = q^n$، قرار دارند، پس $M|S_t(\mathbf{u})| \leq q^n$. ■

مثال ۱۶.۶ با انتخاب $q = 2$ و $t = 1$ از قضیه ۱۵.۶ نتیجه می‌شود $M \leq 2^n/(1+n)$؛ پس با توجه به صحیح بودن M داریم $M \leq \lfloor 2^n/(1+n) \rfloor$. از این‌رو برای $n = 1, 2, 3, 4, 5, 6, 7, \ldots$ رابطه $M \leq 1, 1, 2, 3, 5, 9, 16, \ldots$ برقرار است.

نتیجه ۱۷.۶ هر $[n,k]$–کد خطی t خطا تصحیح کننده C در شرط زیر صدق می‌کند

$$\sum_{i=0}^{t} \binom{n}{i}(q-1)^i \leq q^{n-k}.$$

اثبات چون $\dim(C) = k$ داریم $M = q^k$. نتیجه از تقسیم طرفین نامساوی قضیه ۱۵.۶ بر q^k به‌دست می‌آید. ■

در یک $[n,k]$–کد خطی C هر کدکلمه n رقم دارد که k رقم آن‌را می‌توان به‌عنوان بخش حامل اطلاعات و بقیه را ارقام کنترل درنظر گرفت. نتیجه ۱۷.۶ یک کران پایین روی تعداد ارقام کنترل لازم برای تصحیح t خطا ارائه می‌دهد:

$$n - k \geq \log_q \left(\sum_{i=0}^{t} \binom{n}{i}(q-1)^i \right).$$

یک کد C کامل است اگر در نامساوی قضیه ۱۵.۶ (معادلاً در نتیجه ۱۷.۶، اگر C خطی باشد) تساوی برقرار باشد. این معادل با این است که کره‌های جدا از هم $S_t(\mathbf{u})$، $\mathbf{u} \in C$، کاملاً $\mathcal{V}$ را پر کنند، به‌قسمی‌که هر کلمه $\mathbf{v}$ به فاصله حداکثر t از دقیقاً یک کدکلمه $\mathbf{u}$ قرار داشته باشد. (یک چنین کره‌چینی کاملی در فضای اقلیدسی $\mathbf{R}^n$، $n > 1$، غیر ممکن است، زیرا همیشه شکاف‌های پر نشده‌ای، بین کره‌ها وجود دارند. بهترین بسته‌بندی ممکن در صفحه واضح و شناخته شده است ولی این مسئله در $\mathbf{R}^3$ تا ۱۹۹۸ حل نشده بود؛ برای حل آن به‌وسیله تامس هلز[1] به تارنمای www.math.las.umich.edu/~ hales مراجعه کنید. برای روابط بین نظریه کدگذاری و کره‌چینی اقلیدسی مراجع [CS 92] و [Th 83] را ببینید.)

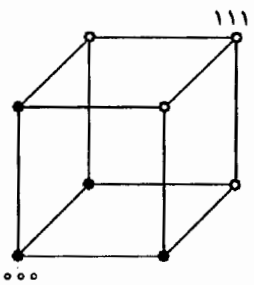
[1] Thomas Hales

تمرین ۵.۶ نشان دهید که یک کد کامل است اگر و فقط اگر به‌ازای عددی چون t، کدگشایی نزدیک‌ترین همسایه تمامی خطاهای به‌وزن حداکثر t را تصحیح کرده ولی هیچ خطای به‌وزن بزرگ‌تر از t را تصحیح نکند.

مثال ۱۸.۶ فرض کنید $\mathcal{C}$ یک کد تکراری دوتایی به‌طول فرد $\mathcal{R}_n$ باشد. این یک کد خطی با $k = ۱$، $q = ۲$، و $t = \lfloor (n-۱)/۲ \rfloor = (n-۱)/۲$ می‌باشد، بنابراین در نتیجه ۱۷.۶ داریم $n - k = n - ۱$. پس با توجه به $q - ۱ = ۱$ و $\binom{n}{i} = \binom{n}{n-i}$ داریم

$$\sum_{i=۰}^{t} \binom{n}{i} (q-۱)^i = \sum_{i=۰}^{t} \binom{n}{i} = \frac{۱}{۲} \sum_{i=۰}^{n} \binom{n}{i} = \frac{۱}{۲} . ۲^n = ۲^{n-۱}.$$

پس کران نتیجه ۱۷.۶ حاصل شده و این کد کامل است. ولی اگر n زوج بوده و یا $q > ۲$ آنگاه $\mathcal{R}_n$ کامل نیست. شکل ۴.۶ دلیل کامل بودن کد دوتایی $\mathcal{R}_۳$ را توضیح می‌دهد. هشت عضو $\mathcal{V} = F_۲^۳$ به دو مجموعه $S_۱(\mathbf{u})$ افراز شده‌اند که $\mathbf{u} \in \{۰۰۰, ۱۱۱\}$. برای هر عدد فرد n افراز مشابه‌ای از $F_۲^n$ به دو مجموعه وجود دارد.



شکل ۴.۶

مثال ۱۹.۶ کد دوتایی همینگ $\mathcal{H}_۷$ یک $[۷, ۴]$–کد خطی است، یعنی این‌که $n = ۷$ و $k = ۴$. چون $q = ۲$ و $t = ۱$، پس

$$\sum_{i=۰}^{t} \binom{n}{i} (q-۱)^i = ۱ + \binom{۷}{۱} = ۸ = q^{n-k}$$

و بنابراین کد کامل است. در فصل ۷ خواهیم دید که این یک کد از خانواده کدهای همینگ دوتایی $\mathcal{H}_n$، $n = ۲^c - ۱$، است که تمامی آنها کامل هستند.

تمرین ۶.۶ کد همینگ دوتایی $\mathcal{H}_۷$ روی کانال BSC، با $P > \frac{۱}{۲}$، به همراه کدگشایی نزدیک‌ترین همسایه Δ به‌کار می‌رود؛ احتمال خطا Pr_E را تعیین کنید. نشان دهید اگر $Q = \overline{P}$ کوچک باشد آنگاه $Pr_E \approx ۲۱Q^۲$.

تمرین ۷.٦ فرض کنید $\mathcal{C}$ کد دوتایی بسط یافته $\overline{\mathcal{H}_۷}$ باشد (تمرین ۴.٦). تعداد کلمات $\mathbf{v} \in \mathcal{V} = F_۲^۸$ را که توسط یک کره $S_t(\mathbf{u})$ پوشانده می‌شوند تعیین کرده و نشان دهید که این کد کامل نیست.

از قضیه ۱۵.٦ برای هر کد دوتایی $\mathcal{C}$ نتیجه می‌شود

$$۲^n \geq M\binom{n}{t} = ۲^{nR}\binom{n}{t}.$$

پس $۲^{n(۱-R)} \geq \binom{n}{t}$، و بنابراین

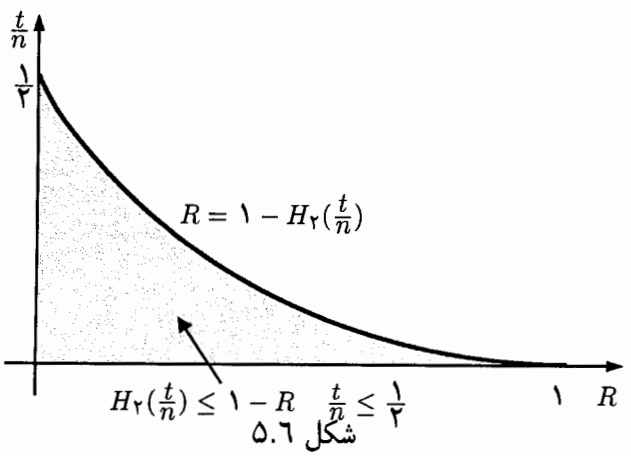
$$۱ - R \geq \frac{۱}{n}\log_۲\binom{n}{t}.$$

اگر از تقریب استرلینگ $n! \sim (n/e)^n\sqrt{۲\pi n}$ (برای نمونه [Fi 83] یا [La 83] را ببینید) برای سه فاکتوریل موجود در $\binom{n}{t} = n!/t!(n-t)!$ استفاده کنیم می‌بینیم که با درنظر گرفتن t/n به‌عنوان یک مقدار ثابت، عبارت سمت راست با افزایش n به $H_۲(t/n)$ میل می‌کند که $H_۲$ تابع دوتایی آنتروپی است (تمرین ۸.٦). در حد نتیجه می‌شود

$$H_۲(\frac{t}{n}) \leq ۱ - R, \qquad (۷.٦)$$

که کران بالای همینگ روی نسبت t/n در کدهای دوتایی با نرخ R و مقادیر بزرگ n می‌باشد. شکل ۵.٦ ناحیه مجاز به‌وسیله این نامساوی را نشان می‌دهد؛ توجه کنید که چون $n \leq d$ و $t = \lfloor (d-۱)/۲ \rfloor$ پس $t/n < ۱/۲$.

مشخص کردن نقاطی از این ناحیه که به کدهای دوتایی مختلفی چون کدهای تکراری، کنترل تساوی و همینگ متناظر می‌شوند تمرین مفیدی است.



شکل ۵.٦

تمرین ۸.٦ ثابت کنید برای مقدار ثابت t/n اگر $n \longrightarrow \infty$ آنگاه $\frac{۱}{n}\log_۲\binom{n}{t}$ به $H_۲(\frac{t}{n})$ میل می‌کند.

۵.۶ کران گیلبرت – ورشامو

برای ماکزیمم نمودن نرخ $R = \frac{1}{n}\log_q M$، در عین حفظ خواص لازم برای خوب بودن یک کد تصحیح خطا، و با فرض داشتن مقادیر q، n و t (یا معادلاً d) علاقمند به پیدا کردن کدهای با بیشترین مقدار $M = |\mathcal{C}|$ هستیم. فرض کنید $A_q(n,d)$ معرف بیشترین تعداد کدکلمه در بین تمامی کدهای q–تایی به‌طول n و می‌نیمم فاصله d باشد که $d \le n$. کران کره‌چینی همینگ (قضیه ۱۵.۶) با نشان دادن رابطه زیر یک کران بالا برای $A_q(n,d)$ ارائه می‌دهد که در آن $t = \lfloor (d-1)/2 \rfloor$:

$$A_q(n,d)\left(1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{t}(q-1)^t\right) \le q^n.$$

مثال ۲۰.۶ اگر $q = 2$ و $d = 3$ آنگاه $t = 1$، از این‌رو همچون مثال ۱۶.۶ ملاحظه می‌کنیم $A_2(n,3) \le \lfloor 2^n/(n+1) \rfloor$. در نتیجه برای $n = 3, 4, 5, 6, 7, \ldots$ داریم $A_2(n,3) \ge 2, 3, 5, 9, 16, \ldots$.

تمرین ۹.۶ متناظر با کران‌های بالای ارائه شده برای $A_2(n,3)$ در مثال ۲۰.۶، کران‌هایی برای $A_3(n,3)$ پیدا کنید. از کران کره‌چینی همینگ چه چیزی راجع به $A_2(n,4)$ و $A_2(n,5)$ به‌دست می‌آید؟

یک بحث مشابه کران پایینی برای $A_q(n,d)$ ایجاد می‌کند؛ این کران گیلبرت – ورشامو[1] است [Va 57]،[Gi 52]:

قضیه ۲۱.۶ اگر $q \ge 2$ و $n \ge d \ge 1$ آنگاه

$$A_q(n,d)\left(1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{d-1}(q-1)^{d-1}\right) \ge q^n.$$

اثبات در میان تمامی کدهای با پارامترهای داده شده q، n و d، فرض کنید $\mathcal{C}$ دارای بیشترین تعداد کدکلمه باشد، پس $M = |\mathcal{C}| = A_q(n,d)$. کره‌های

$$S_{d-1}(\mathbf{u}) = \{\mathbf{v} \in \mathcal{V} \mid d(\mathbf{u},\mathbf{v}) \le d-1\},$$

که $\mathbf{u} \in \mathcal{C}$، فضای $\mathcal{C}$ را می‌پوشانند زیرا اگر کلمه‌ای چون $\mathbf{v} \in \mathcal{V}$ در هیچ کره $S_{d-1}(\mathbf{u})$ نباشد آنگاه برای هر $\mathbf{u} \in \mathcal{C}$ رابطه $d(\mathbf{u},\mathbf{v}) \ge d$ برقرار است، از این‌رو کد $\mathcal{C}' = \mathcal{C} \cup \{\mathbf{v}\}$ با مقادیر n، q و d دارای تعداد کدکلمه بیشتری در مقایسه با $\mathcal{C}$ بوده که در تناقض با

[1] Gilbert - Varshamov bound

انتخاب C است. بنابر بحث به‌کار رفته در اثبات (۶.۶) هر یک از کرات $S_{d-1}(\mathbf{u})$ حاوی $\sum_{i=0}^{d-1}\binom{n}{i}(q-1)^i$ بردار است؛ تمامی q^n عضو $\mathcal{V}$ در بین این بردارها قرار داشته و بنابراین حکم ثابت می‌شود. ■

مثال ۲۲.۶ اگر $q=2$ و $d=3$ (پس $t=1$)، آنگاه برای هر $n\geq 3$ از قضیه ۲۱.۶ نتیجه می‌شود

$$A_2(n,3)\left(1+n+\frac{n(n-1)}{2}\right)\geq 2^n.$$

پس $A_2(n,3)\geq 2^{n+1}/(n^2+n+2)$. چون $A_q(n,d)$ عدد صحیح است پس

$$A_2(n,3)\geq \lceil 2^{n+1}/(n^2+n+2)\rceil.$$

از این نامساوی برای $n=3,4,5,6,7,\ldots$ نتیجه می‌شود $A_2(n,3)\leq 2,2,2,3,5,\ldots$. اگر این کران‌های پایین را با کران‌های بالای ارائه شده در مثال ۲۰.۶ مقایسه کنیم می‌بینیم $A_2(3,3)=2$. به‌عنوان مثال کد دوتایی تکراری $\mathcal{R}_3$ این کران را به‌دست می‌آورد. برای $n=4$ داریم $3\leq A_2(4,3)\leq 2$، پس $A_2(4,3)\in\{2,3\}$.

تمرین ۱۰.۶ نشان دهید $A_2(4,3)=2$ و یک کد با این کران ارائه دهید.

تمرین ۱۱.۶ یک کران پایین برای $A_3(n,3)$ پیدا کنید.

برای اکثر مقادیر q، n و d یک فاصله خیلی زیادی بین کران‌های بالا و پایین ارائه شده برای $A_q(n,d)$ وجود دارد و پیدا کردن مقدار دقیق آن می‌تواند مسئله مشکلی باشد؛ در واقع برای بسیاری از حالات این مقدار هنوز معلوم نیست. در حالت‌های مشخصی وجود یک کد خاص مقدار آن‌را تعیین می‌کند. به‌عنوان مثال برای $q=2$، $d=3$ و $n=7$ کد همینگ $\mathcal{H}_7$ کران بالای $M\leq 16$ را به‌دست می‌آورد، پس $A_2(7,3)=16$. در بخش ۴.۷ خواهیم دید که اگر n به‌فرم 2^c-1 باشد آنگاه $A_2(n,3)$ کران بالای 2^{n-c} را به‌دست می‌آورد.

در حالت دوتایی، قضیه ۲۱.۶ شکل زیر را دارد

$$A_2(n,d)\left(1+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{d-1}\right)\geq 2^n.$$

حال از تمرین ۷.۵ برای $Q<\frac{1}{2}$ نتیجه می‌شود

$$\sum_{i\leq nQ}\binom{n}{i}\leq 2^{nH_2(Q)}.$$

از این‌رو برای $d \leq \lfloor n/2 \rfloor$ داریم

$$\log_2 A_2(n,d) \geq n\left(1 - H_2\left(\frac{d-1}{n}\right)\right).$$

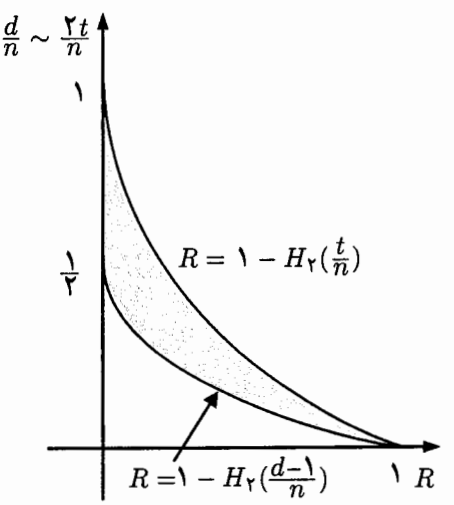
چون نرخ یک کد دوتایی برابر $R = \frac{1}{n}\log_2 M$ است پس اگر $d \leq \lfloor n/2 \rfloor$ آنگاه یک کد با پارامترهای n، q و R موجود است به‌قسمی‌که

$$R \geq 1 - H_2\left(\frac{d-1}{n}\right).$$

می‌توان این کران پایین را با کران جانبی بالای همینگ

$$R \leq 1 - H_2\left(\frac{t}{n}\right),$$

که در بخش ۴.۶ ثابت شد مقایسه کرد. شکل ۶.۶ ناحیه تعریف شده به‌وسیله این دو کران روی R را نشان می‌دهد.



شکل ۶.۶

۶.۶ ماتریس‌های هادامارد و کدها

ساختارهای ریاضی بسیاری را می‌توان برای ساخت کدها به‌کار برد. کلاس جالبی از کدها از روی ماتریس‌های هادامارد ساخته می‌شوند. ابتدا خواص مقدماتی این ماتریس‌ها را بررسی می‌کنیم (برای جزییات بیشتر به [Ha 67، MS 77] مراجعه کنید).

هادامارد علاقمند به دانستن این بود که برای عدد طبیعی مفروض n مقدار دترمینان یک ماتریس $n \times n$ حقیقی $H = (h_{ij})$ تا چه حد می‌تواند بزرگ باشد. برای معنی‌دار بودن موضوع باید کرانی روی درایه‌های H قرار داد و بدون از دست دادن کلیت مسئله می‌توان فرض کرد $|h_{ij}| \leq ۱$. تحت این شرایط هادامارد ثابت کرد $|\det H| \leq n^{n/۲}$، و تساوی برقرار است اگر و فقط اگر

(الف) $h_{ij} = \pm ۱$،

(ب) سطرهای متمایز H متعامد باشند، یعنی اگر $i \neq j$ آنگاه $\mathbf{r}_i.\mathbf{r}_j = ۰$.

یک ماتریس $n \times n$ H که واجد دو خاصیت (الف) و (ب) باشد یک ماتریس هادامارد از مرتبه n نامیده می‌شود. چون از (الف) نتیجه می‌شود $\mathbf{r}_i.\mathbf{r}_i = n$، $۱ \leq i \leq n$، پس HH^T ماتریس قطری زیر است که در آن H^T معرف ترانهاده H بوده، و I_n ماتریس $n \times n$ واحد است:

$$HH^T = \begin{pmatrix} n & ۰ & \cdots & ۰ \\ ۰ & n & \cdots & ۰ \\ \vdots & \vdots & \ddots & \vdots \\ ۰ & ۰ & \cdots & n \end{pmatrix} = nI_n; \qquad (۸.٦)$$

چون $\det H^T = \det H$، از (۸.٦) نتیجه می‌شود

$$(\det H)^۲ = \det(nI_n) = n^n,$$

از این‌رو $|\det H| = n^{n/۲}$. پس تمامی ماتریس‌های هادامارد کران بالای هادامارد را به‌دست می‌آورند. عکس موضوع را که سخت‌تر بوده و در این‌جا به آن نیازی نیست حذف می‌کنیم.

به دلایل زیبایی و چاپی، از نماد '$-$' برای نشان دادن درایه '-۱' در یک ماتریس هادامارد استفاده می‌کنیم.

مثال ۲۳.٦ ماتریس‌های $H = (۱)$ و $H = \begin{pmatrix} ۱ & ۱ \\ ۱ & - \end{pmatrix}$ به‌ترتیب ماتریس‌های هادامارد از مرتبه ۱ و ۲ بوده که برای اولی $|\det H| = ۱$ و برای دومی $|\det H| = ۲$.

تمرین ۱۲.٦ تمامی ماتریس‌های هادامارد از مرتبه ۱ و ۲ را پیدا کنید.

نتیجه ساده زیر امکان ساخت ماتریس‌های هادامارد بزرگ از روی ماتریس‌های کوچک را فراهم می‌کند.

لم ۲۴.٦ فرض کنید H یک ماتریس هادامارد از مرتبه n بوده و

$$H' = \begin{pmatrix} H & H \\ H & -H \end{pmatrix}.$$

در این‌صورت H' یک ماتریس هادامارد از مرتبه $۲n$ است.

تمرین ۱۳.۶ لم ۲۴.۶ را ثابت کنید.

نتیجه ۲۵.۶ برای هر عدد $m \geq 0$ یک ماتریس هادامارد از مرتبه 2^m وجود دارد.

اثبات با $H = (1)$ شروع کرده و لم ۲۴.۶ را m بار به‌کار ببرید. ■

مثال ۲۶.۶ ماتریس‌های هادامارد از مرتبه 2^m که به این طریق به‌دست می‌آیند ماتریس‌های سیلوستر[1] نامیده می‌شوند. به‌عنوان مثال، برای $m = 1$ ماتریس $H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ و برای $m = 2$ ماتریس زیر حاصل می‌شود

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

با این حال ماتریس‌های هادامارد برای هر مرتبه‌ای موجود نیستند. به‌عنوان مثال، یک ماتریس هادامارد از مرتبه فرد $n > 1$ وجود ندارد.

لم ۲۷.۶ اگر H یک ماتریس هادامارد از مرتبه $n > 1$ باشد آنگاه n زوج است.

اثبات از تعامد سطرهای متمایز $\mathbf{r}_i$ و $\mathbf{r}_j$ نتیجه می‌شود $h_{i1}h_{j1} + \cdots + h_{in}h_{jn} = 0$. چون $h_{ik}h_{jk} = \pm 1$، پس n باید زوج باشد. ■

با قدری کار بیشتر محدودیت قوی‌تر زیر روی مرتبه یک ماتریس هادامارد حاصل می‌شود.

لم ۲۸.۶ اگر یک ماتریس هادامارد از مرتبه $n > 2$ موجود باشد آنگاه n مضربی از ۴ است.

اثبات ضرب هر ستون H در -1 خاصیت هادامارد بودن را حفظ می‌کند، از این‌رو می‌توان فرض کرد که درایه‌های سطر اول همگی ۱ هستند. هر سطر $\mathbf{r}_i$، $i \neq 1$، بر $\mathbf{r}_1$ عمود است و بنابراین $n/2$ مؤلفه‌های آن ۱ و بقیه -1 می‌باشند. با اعمال جایگشت (که خاصیت هادامارد بودن را حفظ می‌کند) می‌توان فرض کرد

$$\mathbf{r}_2 = (1 \quad 1 \quad \cdots \quad 1 \quad -1 \quad -1 \quad \cdots \quad -1).$$

حال فرض کنید که اولین و آخرین $n/2$ درایه $\mathbf{r}_3$ به‌ترتیب حاوی u و v مؤلفه برابر با ۱ باشند. در این‌صورت

$$0 = \mathbf{r}_1.\mathbf{r}_3 = u - \left(\frac{n}{2} - u\right) + v - \left(\frac{n}{2} - v\right) = 2u + 2v - n$$

و

[1] Sylvester matrices

$$\circ = \mathbf{r}_۲.\mathbf{r}_۳ = u - \left(\frac{n}{۲} - u\right) - v + \left(\frac{n}{۲} - v\right) = ۲u - ۲v,$$

پس $u = v$ و از این‌رو $n = ۲u + ۲v = ۴u$. ■

حدس زده می‌شود که عکس این قضیه نیز درست باشد، یعنی این‌که برای هر عدد طبیعی n که مضربی از ۴ باشد یک ماتریس هادامارد از مرتبه n موجود است. این هنوز یک مسئله باز است. ارتباط ماتریس‌های هادامارد و نظریه کدگذاری در نتیجه زیر قرار دارد.

قضیه ۲۹.۶ هر ماتریس هادامارد H از مرتبه n منجر به یک کد دوتایی به‌طول n با $M = ۲n$ کدکلمه و می‌نیمم فاصله $d = n/۲$ می‌شود.

اثبات ابتدا از n سطر ماتریس H تعداد $۲n$ بردار $\pm\mathbf{r}_۱, \pm\mathbf{r}_۲, \ldots, \pm\mathbf{r}_n$ در $\mathbf{R}^n$ تشکیل می‌دهیم. تعامد سطرهای H ایجاب می‌کند که این بردارها متمایز باشند. با تبدیل هر درایه ۱− به ۰ تعداد $۲n$ بردار با درایه‌های ۰ و ۱ به‌دست می‌آوریم؛ می‌توان این بردارها را اعضای $\mathcal{V} = F_۲^n$ درنظر گرفت، بنابراین این بردارها تشکیل یک کد دوتایی $\mathcal{C}$ می‌دهند. براساس ساخت این بردارها می‌توان آنها را به‌صورت $\mathbf{u}_۱, \overline{\mathbf{u}}_۱, \ldots, \mathbf{u}_n, \overline{\mathbf{u}}_n$ درنظر گرفت که $\overline{\mathbf{u}} = \mathbf{1} - \mathbf{u}$. برای هر i رابطه $d(\mathbf{u}_i, \overline{\mathbf{u}}_i) = n$ برقرار است و از خاصیت (ب) نتیجه می‌شود هر زوج متمایز از دو کدکلمه دیگر به فاصله $n/۲$ از یکدیگر قرار دارند، بنابراین $\mathcal{C}$ دارای می‌نیمم فاصله $d = n/۲$ است. ■

تمرین ۱۴.۶ تمامی کدکلمه‌های به‌دست آمده با روش فوق از روی ماتریس هادامارد مثال ۲۶.۶ را پیدا کنید. آیا آنها تشکیل یک کد خطی می‌دهند؟

هر کد ساخته شده به روش قضیه ۲۹.۶ یک کد هادامارد به‌طول n نامیده می‌شود. یک کد این‌چنینی به‌طول ۳۲ به‌منظور ارسال تصویر از فضاپیمای مارینر ۱۹۶۹ به‌کار رفت.

تمرین ۱۵.۶ یک ماتریس هادامارد از مرتبه ۸، و بنابراین یک کد هادامارد به‌طول ۸ بسازید. نرخ آن را تعیین کنید. چند خطا را تصحیح و چند خطا را کشف می‌کند؟

اگر n توانی از ۲ نباشد آنگاه $۲n$ نیز این‌گونه خواهد بود. از این‌رو یک کد هادامارد به‌طول یک چنین مقداری از n خطی نیست. نرخ ارسال هر کد هادامارد به‌طول n برابر است با

$$R = \frac{\log_۲(۲n)}{n} = \frac{۱ + \log_۲ n}{n}$$

که با افزایش n به ۰ میل می‌کند. تعداد خطاهای تصحیح شده (اگر $n > ۲$) بنابر قضیه‌های ۱۰.۶ و ۲۹.۶ برابر است با

$$t = \lfloor \frac{d - ۱}{۲} \rfloor = \lfloor \frac{n - ۲}{۴} \rfloor = \frac{n}{۴} - ۱;$$

از این‌رو نسبت خطاهای تصحیح شده برابر

$$\frac{t}{n} = \frac{۱}{۴} - \frac{۱}{n}$$

است که با افزایش n به ۱/۴ میل می‌کند.

۷.۶ تمرینات تکمیلی

تمرین ۱۶.۶ یک چندجمله‌ای تحویل‌ناپذیر درجه ۳ روی $\mathbf{Z}_۲$ پیدا کرده و آن‌را برای ساخت یک میدان هشت عضوی $F_۸$ به‌کار ببرید. نشان دهید که دقیقاً دو چندجمله‌ای با این خاصیت وجود دارند و این‌که میدان‌های به‌دست آمده ایزومورف هستند.

تمرین ۱۷.۶ نشان دهید که برای هر عدد اول $p = ۳ \bmod ۴$ چندجمله‌ای $f(x) = x^۲ + ۱$ روی $\mathbf{Z}_p$ تحویل‌ناپذیر است. از این در ساخت یک میدان F_q از مرتبه $q = p^۲$ استفاده کنید. برای کدام‌یک از اعداد اول p چندجمله‌ای $x^۲ + x + ۱$ منجر به میدان از مرتبه $q = p^۲$ می‌شود؟

تمرین ۱۸.۶ کران یکانی[۱] را ثابت کنید: اگر یک کد روی F_q به‌طول n، می‌نیمم فاصله d و اندازه M کدکلمه باشد آنگاه $\log_q M \leq n - d + ۱$. (راهنمایی: این کد را $d - ۱$ بار پنجر کنید.) چند مثال از کدهایی که این کران را به‌دست می‌آورند ارائه دهید؛ این کدها را جدایی‌پذیر با بیشترین فاصله[۲] می‌نامند.

تمرین ۱۹.۶ اعداد زیر را محاسبه و تجزیه کنید

$$۱ + \binom{۲۳}{۱} + \binom{۲۳}{۲} + \binom{۲۳}{۳}, \qquad ۱ + \binom{۱۱}{۱}.۲ + \binom{۱۱}{۲}.۲^۲.$$

این اعداد چه چیزی را در رابطه با احتمال وجود کدهای کامل بیان می‌کنند؟

تمرین ۲۰.۶ نشان دهید اگر برای $i = ۱, ۲$ کد C_i یک (n, M_i, d_i)- کد در $\mathcal{V} = F_q^n$ باشد آنگاه

$$C_۱ \oplus C_۲ = \{(\mathbf{x}, \mathbf{y}) \in \mathcal{V} \oplus \mathcal{V} \mid \mathbf{x} \in C_۱,\ \mathbf{y} \in C_۲\}$$

یک $(۲n, M_۱M_۲, d)$- کد است که $d = \min\{d_۱, d_۲\}$. نیز نشان دهید

$$C_۱ * C_۲ = \{(\mathbf{x}, \mathbf{x} + \mathbf{y}) \in \mathcal{V} \oplus \mathcal{V} \mid \mathbf{x} \in C_۱,\ \mathbf{y} \in C_۲\}$$

یک $(۲n, M_۱M_۲, d')$- کد است که $d' = \min(۲d_۱, d_۲)$. نشان دهید اگر $C_۱$ و $C_۱$ خطی بوده و از بعد $k_۱$ و $k_۲$ باشند، آنگاه $C_۱ \oplus C_۲$ و $C_۱ * C_۲$ خطی بوده و بعد آنها $k_۱ + k_۲$ است.

---

[۱] Singleton bound [۲] Maximum distance separable codes

تمرین ۲۱.۶ اگر از خطوط رابط صرف‌نظر کنیم آنگاه یک شماره استاندارد بین‌المللی کتاب (ISBN) یک کدکلمه $w = a_1 \dots a_{10}$ به‌طول ۱۰ روی $\mathbf{Z}_{11} = \{0, 1, \dots, 9, X\}$ است که 'X' معرف ۱۰ می‌باشد. ارقام a_1،...،a_9 ارقام اطلاعات هستند که معرف کشور، انتشارات، و مؤلفه‌های دیگری می‌باشند درحالی‌که a_{10} یک رقم کنترلی تعریف شده با $a_1 + 2a_2 + \cdots + 10a_{10} = 0 \bmod 11$ است. نشان دهید این کد می‌تواند هر یک رقم نادرست و نیز جابه‌جایی هر دو رقمی را کشف کند (اینها از رایج‌ترین خطاهای بشر هستند). کدام‌یک از اعداد زیر ISBN معتبر هستند؟

۳ – ۵۴۰ – ۷۶۱۹۷ – ۷, ۳ – ۵۴۰ – ۷۶۱۷۹ – ۷, ۳ – ۵۴۱ – ۷۶۱۹۷ – ۷.

فصل ۷

کدهای خطی

وضعیت من و سبب آن‌را به‌درستی گزارش کن. (Hamlet)

در فصل ٦ چند نمونه از کدهای خطی را ملاحظه کردیم. پیش از این در لم ۸.٦ یک نقطه قوت کار کردن با کدهای خطی را دیده‌ایم که در واقع سهولت محاسبه می‌نیمم فاصله آنها در مقایسه با محاسبه این پارامتر در کدهای غیر خطی می‌باشد. در این فصل کدهای خطی را با جزییات بیشتری مطالعه خواهیم کرد؛ به نقاط قوت دیگری از این کدها با استفاده از جبر خطی مقدماتی و نظریه ماتریس‌ها اشاره کرده که شامل حتی یک روش ساده‌تری برای محاسبه می‌نیمم فاصله کد نیز می‌باشد. پیشینه نظری لازم حاوی مباحثی چون استقلال خطی، بعد، و عملیات سطری و ستونی می‌باشد. اینها معمولاً در یک درس جبرخطی سال اول دانشگاه پوشانده می‌شوند؛ اگرچه دروس این‌چنینی اغلب تمرکز روی فضاهای برداری و ماتریس‌های حقیقی یا مختلط دارند تمامی نتایج مهم و روش‌هایی که ما لازم داریم به یک میدان دلخواه، شامل میدان‌های متناهی، تعمیم داده می‌شوند. در طول این فصل فرض براین است که الفبای F میدان متناهی F_q است که q توانی از یک عدد اول می‌باشد.

۱.۷ توصیف ماتریسی کدهای خطی

می‌توان یک کد خطی $C \subseteq \mathcal{V} = F^n$ را با ارائه یک مبنای $\mathbf{u}_1, \ldots, \mathbf{u}_k$ برای C مشخص نمود به‌قسمی‌که کدکلمات $\mathbf{u} \in C$ دقیقاً ترکیبات خطی $a_1\mathbf{u}_1 + \cdots + a_k\mathbf{u}_k$ باشند که $a_i \in F$. بر این اساس به‌جای لیست کردن تمامی $M = q^k$ بردار در C تنها ملزم به

لیست کردن k بردار خواهیم بود که $k = \dim(\mathcal{C})$. یک روش مفید در مشخص نمودن یک مبنا ارائه یک ماتریس مولد G برای $\mathcal{C}$ است؛ این ماتریس دارای k سطر و n ستون بوده که هر سطر آن یک عضو از بردارهای مبنای $\mathcal{C}$ است. (توجه کنید که $\mathcal{C}$ ماتریس G را به‌طور منحصربه‌فرد تعیین نمی‌کند، یک زیرفضا ممکن است مبناهای بسیاری داشته و بردارهای یک مبنا می‌توانند با ترتیب‌های مختلفی نوشته شوند.)

مثال ۱.۷ کد تکراری $\mathcal{R}_n$ روی F دارای مبنای یک برداری $\mathbf{u}_1 = 11\cdots1$ است، بنابراین ماتریس مولد

$$G = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}$$

را دارد که متشکل از یک سطر و n ستون است.

مثال ۲.۷ کد کنترل تساوی $\mathcal{P}_n$ روی F دارای مبنای $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ است که برحسب مبنای استاندارد $\mathbf{e}_1, \ldots, \mathbf{e}_n$ از F^n بردار $\mathbf{u}_i$ دارای نمایش $\mathbf{u}_i = \mathbf{e}_i - \mathbf{e}_n$ است. بنابراین $\mathcal{P}_n$ دارای ماتریس مولد G است

$$G = \begin{pmatrix} 1 & & & -1 \\ & 1 & & -1 \\ & & \ddots & \vdots \\ & & 1 & -1 \end{pmatrix}$$

که $n-1$ سطر و n ستون داشته و در آن درایه‌های حذف‌شده همگی ۰ هستند.

مثال ۳.۷ در مثال ۵.۶، برای کد همینگ $\mathcal{H}_7$ مبنای متشکل از $\mathbf{u}_1 = 1111000$، $\mathbf{u}_2 = 1001100$، $\mathbf{u}_3 = 0101010$، $\mathbf{u}_4 = 1101001$ ارائه شد. این کد بنابراین دارای ماتریس مولد G است

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

اگر یک کد $\mathcal{C}$ دارای بعد k باشد، آنگاه می‌توان فضای k بعدی $\mathcal{A} = F^k$ را به‌عنوان یک منبع درنظر گرفت که با تبدیل خطی $\mathcal{A} \longrightarrow \mathcal{C} \subseteq \mathcal{V} = F^n$، معرفی شده با ماتریس مولد G از کد $\mathcal{C}$، کدگذاری می‌شود. مشخصاً هر کلمه $\mathbf{a} = a_1 \cdots a_k \in \mathcal{A}$ با کدکلمه $\mathbf{u} = \mathbf{a}G \in \mathcal{C}$ کدگذاری می‌شود. از این‌رو عمل کدگذاری در واقع ضرب کردن در یک ماتریس ثابت است که به‌سادگی اجرا می‌شود.

مثال ۴.۷ کد تکراری $\mathcal{R}_n$ دارای بعد $k = ۱$ است، پس $\mathcal{A} = F^۱ = F$. هر $\mathbf{a} \in \mathcal{A}$ با $\mathbf{u} = \mathbf{a}G = a \cdots a$ کدگذاری می‌شود.

مثال ۵.۷ اگر $\mathcal{C} = \mathcal{P}_n$ آنگاه $k = n - ۱$، بنابراین $\mathcal{A} = F^{n-۱}$. هر $\mathbf{a} = a_۱ \cdots a_{n-۱} \in \mathcal{A}$ با $\mathbf{u} = \mathbf{a}G = a_۱ \cdots a_{n-۱} a_n$ کد می‌شود که $a_n = -(a_۱ + \cdots + a_{n-۱})$، پس $\sum_i a_i = ۰$.

مثال ۶.۷ اگر $\mathcal{C} = \mathcal{H}_۷$ آنگاه $n = ۷$ و $k = ۴$، پس $\mathcal{A} = F_۲^۴$. هر $\mathbf{a} = a_۱ \ldots a_۴ \in \mathcal{A}$ با $\mathbf{u} = \mathbf{a}G \in \mathcal{H}_۷$ کد می‌شود. به‌عنوان نمونه در مثال ۵.۶، $\mathbf{a} = ۰۱۱۰$ با $\mathbf{u} = ۱۱۰۰۱۱۰$ کدگذاری شد

$$\mathbf{u} = \mathbf{a}G = (۰\ ۱\ ۱\ ۰) \begin{pmatrix} ۱ & ۱ & ۱ & ۰ & ۰ & ۰ & ۰ \\ ۱ & ۰ & ۰ & ۱ & ۱ & ۰ & ۰ \\ ۰ & ۱ & ۰ & ۱ & ۰ & ۱ & ۰ \\ ۱ & ۱ & ۰ & ۱ & ۰ & ۰ & ۱ \end{pmatrix} = (۱\ ۱\ ۰\ ۰\ ۱\ ۱\ ۰).$$

اگر G یک ماتریس مولد مفروض برای کد $\mathcal{C}$ باشد، آنگاه تعیین این‌که آیا یک بردار $\mathbf{v} \in \mathcal{V}$ در $\mathcal{C}$ قرار دارد و این‌که اگر در $\mathcal{C}$ نیست کدام عضو $\mathbf{u} \in \mathcal{C}$ به $\mathbf{v}$ کمترین فاصله را دارد می‌تواند عملی کاملاً کسل کننده باشد. برای ساده‌تر کردن این مسئله یک توصیف ماتریسی دیگر از $\mathcal{C}$ جستجو می‌کنیم. یک راه مؤثر در انجام این امر ارائه یک دستگاه متشکل از $n - k$ معادله خطی برای معرفی اعضای $\mathcal{C}$ است، به‌طوری‌که $\mathbf{v} \in \mathcal{V}$ در $\mathcal{C}$ قرار دارد اگر و فقط اگر درایه‌های آن در این دستگاه صدق کنند.

مثال ۷.۷ کد تکراری $\mathcal{R}_n$ متشکل است از بردارهای $\mathbf{v} = v_۱ \ldots v_n \in \mathcal{V}$ که در شرط $v_۱ = \cdots = v_n$ صدق کنند این شرط را می‌توان به‌عنوان $n - k = n - ۱$ معادله خطی هم‌زمان $v_i - v_n = ۰$، $۱ \le i \le n - ۱$، درنظر گرفت.

مثال ۸.۷ کد کنترل تساوی $\mathcal{P}_n$ $(n - k = ۱)$ زیرفضای $\mathcal{V}$ است که با معادله خطی $v_۱ + \cdots + v_n = ۰$ تعریف می‌شود.

مثال ۹.۷ کد همینگ $\mathcal{H}_۷$ متشکل از بردارهای $\mathcal{V} = F_۲^۷$ است که در معادلات زیر صدق می‌کنند:

$$v_۴ + v_۵ + v_۶ + v_۷ = ۰,$$
$$v_۲ + v_۳ + v_۶ + v_۷ = ۰,$$
$$v_۱ + v_۳ + v_۵ + v_۷ = ۰.$$

در حالت کلی c معادله مستقل خطی یک زیرفضای $n-c$ بعدی از $\mathcal{V}$ تعریف می‌کنند، از این‌رو برای مشخص نمودن $\mathcal{C}$ به $c=n-k$ معادله مستقل نیاز داریم. اینها معادلات کنترل تساوی نامیده می‌شوند و ماتریس ضرایب آنها، H، که n ستون و $n-k$ سطر مستقل دارد ماتریس کنترل تساوی برای $\mathcal{C}$ نامیده می‌شود. معادلات خطی را می‌توان به‌فرم $\mathbf{v}H^T=\mathbf{0}$ نمایش داد که H^T ترانهاده ماتریس H است، بنابراین آزمون مفید و کارای زیر را برای اعضای $\mathcal{C}$ داریم:

لم ۱۰.۷ فرض کنید $\mathcal{C}$ یک کد خطی در $\mathcal{V}$ بوده و ماتریس کنترل تساوی آن H باشد. آنگاه کلمه $\mathbf{v}\in\mathcal{V}$ در $\mathcal{C}$ قرار دارد اگر و فقط اگر $\mathbf{v}H^T=\mathbf{0}$.

مثال ۱۱.۷ با به‌کار بردن معادلات $v_i-v_n=0$ برای $1\le i\le n-1$، می‌بینیم ماتریس H

$$H=\begin{pmatrix} 1 & & & -1\\ & 1 & & -1\\ & & \ddots & \vdots\\ & & 1 & -1 \end{pmatrix},$$

که دارای $n-1$ سطر و n ستون است یک ماتریس کنترل تساوی برای $\mathcal{R}_n$ است.

مثال ۱۲.۷ معادله $v_1+\cdots+v_n=0$ نشان می‌دهد که می‌توانیم ماتریس

$$H=\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix},$$

را که با یک سطر و n ستون می‌باشد به‌عنوان ماتریس کنترل تساوی برای $\mathcal{P}_n$ درنظر بگیریم.

مثال ۱۳.۷ سه معادله خطی داده شده در مثال ۹.۷ برای $\mathcal{H}_7$، ماتریس کنترل تساوی H را فراهم می‌کنند

$$H=\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1\\ 0 & 1 & 1 & 0 & 0 & 1 & 1\\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

تمرین ۱.۷ اگر $\mathcal{C}$ یک کد خطی با ماتریس مولد G و ماتریس کنترل تساوی H باشد، یک ماتریس مولد $\overline{G}$ و ماتریس کنترل تساوی $\overline{H}$ برای کد بسط یافته $\overline{\mathcal{C}}$ پیدا کنید.

تمرین ۲.۷ اگر $\mathcal{C}_1$ و $\mathcal{C}_2$ کدهای خطی در $\mathcal{V}$ باشند و به‌ترتیب دارای ماتریس‌های مولد G_1 و G_2 و ماتریس‌های کنترل تساوی H_1 و H_2 باشند چگونگی پیدا کردن یک ماتریس مولد برای $\mathcal{C}_1+\mathcal{C}_2$ و یک ماتریس کنترل تساوی برای $\mathcal{C}_1\cap\mathcal{C}_2$ را توضیح دهید.

می‌توان H را به‌عنوان ماتریس یک تبدیل خطی $h : \mathcal{V} \longrightarrow \mathcal{W} = F^{n-k}$ درنظر گرفت که هر بردار $\mathbf{v} \in \mathcal{V}$ را به $h(\mathbf{v}) = \mathbf{v}H^T \in \mathcal{W}$ می‌نگارد، از این‌رو لم ۱۰.۷ می‌گوید $\mathcal{C}$ هسته h، $\ker(h)$، است یعنی مجموعه بردارهایی که به $\mathbf{0}$ نگاشته می‌شوند. تصویر h، $\mathrm{im}(h)$، زیرفضایی از $\mathcal{W}$ است که توسط ستون‌های H تولید می‌شود. قضیه بعد $\dim(\mathcal{V}) = \dim(\ker(h)) + \dim(\mathrm{im}(h))$ ایجاب می‌کند $\dim(\mathrm{im}(h)) = n - k$؛ پس h فضای $\mathcal{V}$ را به‌روی $\mathcal{W}$ می‌نگارد. سطرهای H، که معرف معادلات خطی مشخص کننده $\mathcal{C}$ هستند، مستقل خطی می‌باشند و بنابراین تشکیل یک مبنا برای یک زیرفضای $\mathcal{D} \subseteq \mathcal{V}$ به بعد $n - k$ می‌دهند؛ این یک کد خطی با ماتریس مولد H می‌باشد که دوگان $\mathcal{C}$ نامیده می‌شود.

کدهای $\mathcal{C}$ و $\mathcal{D}$ با مفهوم تعامد با هم ارتباط دارند. همچون فضای اقلیدسی، ضرب اسکالر دو بردار $\mathbf{u} = u_1 \dots u_n$ و $\mathbf{v} = v_1 \dots v_n$ از اعضای $\mathcal{V} = F^n$ را با

$$\mathbf{u}.\mathbf{v} = u_1 v_1 + \cdots + u_n v_n \in F \qquad (۱.۷)$$

تعریف می‌کنیم. این ضرب نسبت به هر دو متغیر خطی است، یعنی این‌که برای هر $a, b \in F$ روابط زیر برقرارند:

$$(a\mathbf{u}_1 + b\mathbf{u}_2).\mathbf{v} = a(\mathbf{u}_1.\mathbf{v}) + b(\mathbf{u}_2.\mathbf{v}), \qquad \mathbf{u}.(a\mathbf{v}_1 + b\mathbf{v}_2) = a(\mathbf{u}.\mathbf{v}_1) + b(\mathbf{u}.\mathbf{v}_2).$$

دو بردار $\mathbf{u}$ و $\mathbf{v}$ را متعامد نامیم هر گاه $\mathbf{u}.\mathbf{v} = 0$. بر خلاف $\mathbf{R}^n$، یک بردار ناصفر می‌تواند بر خودش عمود باشد. برای نمونه اگر $\mathbf{u} = \mathbf{e}_1 + \mathbf{e}_2$ آنگاه $\mathbf{u}.\mathbf{u} = 1^2 + 1^2 = 2$، از این‌رو $\mathbf{u}.\mathbf{u} = 0$ اگر $q = 2^e$.

معادله $\mathbf{v}H^T = 0$ که $\mathcal{C}$ را تعریف می‌کند می‌تواند به این شکل تفسیر شود که $\mathcal{C}$ متشکل از تمامی بردارهایی است که بر هر سطر از ماتریس H عمود باشند، یا معادلاً بر تمامی اعضای $\mathcal{D}$ عمود باشند. پس $\mathcal{C}$ کد متعامد $\mathcal{D}^\perp = \{\mathbf{v} \in \mathcal{V} \mid \mathbf{v}.\mathbf{w} = 0, \quad \mathbf{w} \in \mathcal{D}\}$ است. با تعویض نقش $\mathcal{C}$ و $\mathcal{D}$ می‌بینیم که

$$\mathcal{D} = \mathcal{C}^\perp = \{\mathbf{w} \in \mathcal{V} \mid \mathbf{v}.\mathbf{w} = 0, \quad \mathbf{v} \in \mathcal{C}\}.$$

بنابراین کدهای خطی به‌صورت زوج‌های دوگانی ظاهر می‌شوند؛ ماتریس مولد یکی از آنها ماتریس کنترل تساوی برای دیگری است. با این حال توجه کنید که بعضی از کدهای خطی خود–دوگان هستند، یعنی $\mathcal{C} = \mathcal{C}^\perp$؛ کد دوتایی تکراری $\mathcal{R}_2$ مثالی ساده از این نوع است.

مثال ۱۴.۷ فرض کنید $q = 2$، $n = 2m$، و فرض کنید $\mathcal{C}$ یک کد خطی با بردارهای مبنای $\mathbf{u}_i = \mathbf{e}_{2i-1} + \mathbf{e}_{2i}$، $1 \le i \le m$، باشد. چون $\mathbf{u}_i.\mathbf{u}_j = 0$ پس $\mathcal{C} \subseteq \mathcal{C}^\perp$. با مقایسه ابعاد دو کد می‌بینیم که $\mathcal{C} = \mathcal{C}^\perp$.

مثال ۱۵.۷ کد تکراری $\mathcal{R}_n$ به‌وسیله $\mathbf{1} = ۱۱\ldots۱$ تولید می‌شود، پس

$$\mathcal{R}_n^{\perp} = \{\mathbf{w} \in \mathcal{V} \mid \mathbf{1}.\mathbf{w} = ۰\} = \{\mathbf{w} \in \mathcal{V} \mid w_۱ + \cdots + w_n = ۰\} = \mathcal{P}_n,$$

و مشابهاً

$$\begin{aligned}\mathcal{P}_n^{\perp} &= \{\mathbf{w} \in \mathcal{V} \mid (\mathbf{e}_i - \mathbf{e}_n).\mathbf{w} = ۰, \quad i = ۱,\ldots,n-۱\} \\ &= \{\mathbf{w} \in \mathcal{V} \mid w_i = w_n, \quad i = ۱,\ldots,n-۱\} \\ &= \mathcal{R}_n.\end{aligned}$$

می‌دانیم که ماتریس

$$\begin{pmatrix} ۱ & & & -۱ \\ & ۱ & & -۱ \\ & & \ddots & \vdots \\ & & ۱ & -۱ \end{pmatrix}$$

یک ماتریس مولد برای $\mathcal{P}_n$ و یک ماتریس کنترل تساوی برای $\mathcal{R}_n$ است، درحالی‌که

$$\begin{pmatrix} ۱ & ۱ & \cdots & ۱ \end{pmatrix}$$

یک ماتریس مولد برای $\mathcal{R}_n$ و یک ماتریس کنترل تساوی برای $\mathcal{P}_n$ است.

مثال ۱۶.۷ کد $\mathcal{H}_۳^{\perp}$ یک $[۷, ۳]$– کد خطی روی $F_۲$ است؛ ماتریس زیر یک ماتریس مولد برای این کد و یک ماتریس کنترل تساوی برای $\mathcal{H}_۳$ است

$$\begin{pmatrix} ۰ & ۰ & ۰ & ۱ & ۱ & ۱ & ۱ \\ ۰ & ۱ & ۱ & ۰ & ۰ & ۱ & ۱ \\ ۱ & ۰ & ۱ & ۰ & ۱ & ۰ & ۱ \end{pmatrix}$$

با درنظر گرفتن ترکیبات خطی این سه سطر می‌بینیم که تمامی هفت عضو ناصفر این کد به وزن ۴ هستند، بنابراین $d = ۴$.

این بخش را با یک محک کلی برای تعیین این‌که چه ماتریس‌هایی ماتریس کنترل تساوی یک کد مفروض هستند خاتمه می‌دهیم. در این گزاره $\mathbf{0}$ معرف ماتریسی است که تمامی درایه‌های آن صفر هستند.

لم ۱۷.۷ فرض کنید $\mathcal{C}$ یک $[n, k]$–کد خطی روی F بوده و G یک ماتریس مولد آن باشد. نیز فرض کنید H یک ماتریس با n ستون و $n-k$ سطر روی F باشد. در این‌صورت H یک ماتریس کنترل تساوی برای $\mathcal{C}$ است اگر و فقط اگر H دارای رتبه $n-k$ بوده و $GH^T = \mathbf{0}$.

اثبات سطرهای H تشکیل $n-k$ بردار در $\mathcal{V}$ می‌دهند، و $GH^T = \mathbf{0}$ اگر و فقط اگر این سطرها به سطرهای G عمود باشند، یا معادلاً در $\mathcal{C}^\perp$ قرار گیرند. حال H دارای رتبه $n-k$ است اگر و فقط اگر سطرهای آن مستقل خطی باشند، یا معادلاً تشکیل یک مبنا برای $\mathcal{C}^\perp$ بدهند. بنابراین H در شرایط داده شده صدق می‌کند اگر و فقط اگر یک ماتریس مولد برای $\mathcal{C}^\perp$ باشد، یعنی این‌که یک ماتریس کنترل تساوی برای $\mathcal{C}$ باشد. ■

۲.۷ معادل بودن کدهای خطی

یک فضای برداری در حالت کلی یک مبنای منحصربه‌فرد ندارد، از این‌رو ماتریس‌های مولد G و کنترل تساوی H یک کد خطی $\mathcal{C}$ در حالت کلی منحصربه‌فرد نیستند. مفید خواهد بود اگر این ماتریس‌ها به‌گونه‌ای انتخاب شوند که تا حد ممکن دارای شکل ساده‌ای باشند، به‌عنوان مثال تعداد زیادی از درایه‌ها صفر باشند تا محاسبات ساده‌تر انجام شود.

سطرهای $\mathbf{r}_1, \ldots, \mathbf{r}_k$ به‌عنوان اعضای $\mathcal{V}$، تشکیل یک مبنا برای $\mathcal{C}$ می‌دهند. عملیات سطری مقدماتی شامل جایگشت روی سطرها، ضرب یک سطر در یک اسکالر ناصفر، و جایگزینی یک سطر $\mathbf{r}_i$ با $\mathbf{r}_i + a\mathbf{r}_j$ می‌باشد که $j \neq i$ و $a \neq 0$. این اعمال می‌توانند مبنای $\mathcal{C}$ را عوض کنند ولی زیرفضای تولید شده $\mathcal{C}$ توسط این سطرها را عوض نمی‌کنند.

با این حال، اگر ستون‌های G را تحت یک جایگشت قرار دهیم ممکن است $\mathcal{C}$ را عوض کنیم، ولی اختلاف کد حاصل با $\mathcal{C}$ تنها در ترتیب سمبل‌ها در کدکلمات است؛ هر دو کد دارای پارامترهای یکسانی چون n، k، d، t، M، R و غیره می‌باشند. پس این دو کد به شکل اساسی متفاوت نیستند. این انگیزه‌ای برای تعریف زیر می‌باشد. دو کد خطی $\mathcal{C}_1$ و $\mathcal{C}_2$ معادل هستند اگر به‌ترتیب دارای ماتریس‌های مولد G_1 و G_2 بوده به‌قسمی‌که اختلاف G_1 و G_2 تنها در عملیات سطری و جایگشت روی ستون‌ها باشد. این بدین معنی است که $\mathcal{C}_2$ از روی $\mathcal{C}_1$ با تجدید آرایش هم‌زمان سمبل‌ها در تمامی اعضای $\mathcal{C}_1$ حاصل شود. به‌صورت غیررسمی، تمایل در این است که $\mathcal{C}_1$ و $\mathcal{C}_2$ به‌عنوان کدهای یکسان منظور شوند اگرچه آنها در حالت کلی از کدکلمه‌های متفاوتی تشکیل می‌شوند. با به‌کار بردن عملیات سطری و جایگشت ستون‌ها می‌توان هر ماتریس مولدی را به‌فرم زیر تبدیل کرد

$$G = (I_k|P) = \begin{pmatrix} 1 & & & * & * & \cdots & * \\ & 1 & & * & * & \cdots & * \\ & & \ddots & \vdots & \vdots & & \vdots \\ & & 1 & * & * & \cdots & * \end{pmatrix}, \qquad (۲.۷)$$

که I_k ماتریس همانی $k \times k$ است و P یک ماتریس با k سطر و $n-k$ ستون است که به‌وسیله ستاره‌ها مشخص شده است. در این‌صورت گوئیم G (یا $\mathcal{C}$) فرم منظم دارد. در این حالت هر $\mathbf{a} = a_1 \ldots a_k \in F^k$ با $\mathbf{u} = \mathbf{a}G = a_1 \ldots a_k a_{k+1} \ldots a_n$ کدگذاری می‌شود که $a_1, \ldots, a_k$

ارقام اطلاعات هستند و $a_{k+1}\dots a_n = \mathbf{a}P$ یک قالب از $n-k$ رقم کنترلی می‌باشد. ارقام اطلاعات کاملاً دلخواه بوده، در حالی که ارقام کنترلی به‌صورت منحصربه‌فردی توسط $\mathbf{a}$ و G تعیین شده و به‌سادگی به‌عنوان سمبل‌هایی در $\mathbf{a}P$ محاسبه می‌شوند.

مثال ۱۸.۷ ماتریس‌های مولد G برای کدهای $\mathcal{R}_n$ و $\mathcal{P}_n$ در بخش ۱.۷ به‌فرم منظم هستند.

مثال ۱۹.۷ ماتریس مولد G_1 برای کد $\mathcal{H}_7$ که در بخش ۱.۷ ارائه شد به‌فرم منظم نیست

$$G_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

با این حال با عمل جایگشت روی ستون‌های G_1 یک ماتریس مولد G_2 برای یک کد معادل حاصل می‌شود که به‌فرم منظم است؛ به‌عنوان نمونه جایگشت $\pi = (1\ 7\ 4\ 5\ 2\ 6\ 3)$ منجر به ماتریس زیر می‌شود

$$G_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

تمرین ۳.۷ ماتریس‌های G_1 و G_2 در مثال ۱۹.۷ دو کد معادل تولید می‌کنند. آیا این دو کد مساوی بوده یا متمایز هستند؟

اگر یک ماتریس مولد منظم $G = (I_k|P)$ برای یک کد خطی $\mathcal{C}$ داشته باشیم آنگاه می‌توانیم یک ماتریس کنترل تساوی

$$H = (-P^T|I_{n-k}) \qquad (۳.۷)$$

برای $\mathcal{C}$ داشته باشیم. این فرم منظم برای یک ماتریس کنترل تساوی است. به کمک لم ۱۷.۷ می‌توان نشان داد که این یک ماتریس کنترل تساوی است. ماتریس نشان داده شده H دارای $n-k$ سطر و n ستون بوده، و حضور ماتریس واحد I_{n-k} ثابت می‌کند که سطرهای H مستقل هستند و از این‌رو رتبه H برابر $n-k$ است؛ سرانجام ملاحظه می‌شود که

$$GH^T = I_k(-P) + PI_{n-k} = -P + P = \mathbf{0}.$$

اگر q توانی از ۲ باشد آنگاه برای هر $a \in F$ داریم $a + a = 2a = 0$؛ از این‌رو $-a = a$ و با حذف علائم منفی H فرم ساده‌تر $H = (P^T|I_{n-k})$ را پیدا می‌کند.

مثال ۲۰.۷ در مثال ۱.۷ یک ماتریس مولد منظم G برای $\mathcal{R}_n$ پیدا کردیم که در آن $P = (۱\ ۱\ \cdots\ ۱)$ دارای $n-۱$ درایه می‌باشد. در نتیجه ماتریس کنترل تساوی H، با $n-۱$ سطر، به فرم زیر حاصل می‌شود

$$H = \begin{pmatrix} -۱ & ۱ & & & \\ -۱ & & ۱ & & \\ \vdots & & & \ddots & \\ -۱ & & & & ۱ \end{pmatrix}.$$

این با ماتریس کنترل تساوی داده شده در مثال ۱۱.۷ یکی نیست، ولی منجر به یک دستگاه معادلات کنترل تساوی معادل برای $\mathcal{R}_n$ می‌شود، که با $-v_۱ + v_i = ۰$، $۲ \le i \le n$، تعریف می‌شوند.

مثال ۲۱.۷ ماتریس مولد کد $\mathcal{P}_n$ که در مثال ۲.۷ داده شده به‌فرم منظم است. چون $P = (-۱\ -۱\ \cdots\ -۱)^T$ دارای $n-۱$ درایه است پس $\mathcal{P}_n$ دارای ماتریس کنترل تساوی H است که n درایه دارد:

$$H = (۱\ ۱\ \cdots\ ۱).$$

مثال ۲۲.۷ در ماتریس مولد منظم داده‌شده در مثال ۱۹.۷ برای $\mathcal{H}_۷$ داریم

$$P = \begin{pmatrix} ۰ & ۱ & ۱ \\ ۱ & ۰ & ۱ \\ ۱ & ۱ & ۰ \\ ۱ & ۱ & ۱ \end{pmatrix}.$$

در این‌جا $q = ۲$، و از این‌رو با حذف علامت‌های منفی ماتریس کنترل تساوی استاندارد H برای $\mathcal{H}_۷$ به‌دست می‌آید

$$H = \begin{pmatrix} ۰ & ۱ & ۱ & ۱ & ۱ & ۰ & ۰ \\ ۱ & ۰ & ۱ & ۱ & ۰ & ۱ & ۰ \\ ۱ & ۱ & ۰ & ۱ & ۰ & ۰ & ۱ \end{pmatrix}.$$

به بیان دقیق، اینها ماتریس‌های مولد و کنترل تساوی برای یک کدی هستند که معادل $\mathcal{H}_۷$ است زیرا از اعمال یک جایگشت روی ماتریس‌های مربوط به $\mathcal{H}_۷$ به‌دست آمده‌اند؛ با این حال همچنان‌که قبلاً گفته شد ما تمایزی بین کدهای معادل قائل نخواهیم بود.

با به‌کار بردن ماتریس‌های مولد منظم یک اثبات دیگر برای کران یکانی (تمرین ۱۸.۶) به‌دست می‌آید:

قضیه ۲۳.۷ اگر C یک کد خطی به‌طول n، بعد k، و می‌نیمم فاصله d باشد، آنگاه

$$d \leq n - k + ۱.$$

اثبات با به‌کار بردن یک کد معادل، می‌توان فرض کرد که C یک ماتریس مولد منظم $G = (I_k|P)$ دارد. در این‌صورت هر سطر G یک کدکلمه ناصفر به وزن حداکثر $n - k + ۱$ است؛ این سطر دقیقاً یک رقم اطلاعات ناصفر (در I_k) داشته، و $n - k$ رقم کنترل (در P) دارد، بنابراین حداکثر $n - k + ۱$ رقم آن ناصفر است. از لم ۸.۶ نتیجه می‌شود $d \leq ۱ + n - k$. ■

مثال ۲۴.۷ کران یکانی توسط $\mathcal{R}_n$ با $k = ۱$ و $d = n$ و توسط $\mathcal{P}_n$ با $k = n - ۱$ و $d = ۲$ به‌دست می‌آید. ولی $\mathcal{H}_۷$ با $d = ۳$ و $n - k + ۱ = ۴$ به این کران نمی‌رسد.

نتیجه ۲۵.۷ یک $[n, k]$– کد t خطا تصحیح کننده حداقل به $۲t$ رقم کنترل نیاز دارد.

اثبات تعداد $n - k$ رقم کنترل وجود دارد، و از قضایای ۲۳.۷ و ۱۰.۶ نتیجه می‌شود $۲t \leq d - ۱ \leq n - k$. ■

مثال ۲۶.۷ کدهای خطی $\mathcal{R}_۳$ و $\mathcal{H}_۷$ هر دو دارای $t = ۱$ هستند؛ تعداد رقم‌های کنترل به‌ترتیب برابر ۲ و ۳ است.

۳.۷ می‌نیمم فاصله کدهای خطی

در این بخش چگونگی به‌دست آوردن می‌نیمم فاصله یک کد خطی از روی ماتریس کنترل تساوی بررسی می‌شود.

قضیه ۲۷.۷ فرض کنید C یک کد خطی با می‌نیمم فاصله d بوده و H یک ماتریس کنترل تساوی برای C باشد. در این‌صورت d برابر است با کمترین تعداد از ستون‌های وابسته خطی در H.

اثبات بنابر لم ۱۰.۷ یک بردار $\mathbf{v} = v_۱ \cdots v_n \in \mathcal{V}$ یک کدکلمه است اگر و فقط اگر $\mathbf{v}H^T = \mathbf{0}$ یا معادلاً $\sum_i v_i \mathbf{c}_i = \mathbf{0}$ که $\mathbf{c}_۱, \ldots, \mathbf{c}_n$ ستون‌های H هستند. اگر $\mathbf{v} \neq \mathbf{0}$ آنگاه این یک رابطه وابستگی خطی بین ستون‌ها است و به‌عکس هر یک‌چنین رابطه‌ای متناظر با یک کدکلمه ناصفر $\mathbf{v}$ می‌باشد؛ تعداد ستون‌های ظاهر شده در این معادله برابر تعداد مؤلفه‌های ناصفر v_i است، که در واقع برابر وزن $\mathbf{v}$ است. پس کمترین تعداد ستون‌های وابسته خطی در H برابر است با کمترین وزن کدکلمات ناصفر، و بنابر لم ۸.۶ این عدد برابر d است. ■

قبل از این‌که نگاهی به چند مثال داشته باشیم اجازه دهید ببینیم بستگی خطی یک یا دو ستون از H به چه معنی است. یک ستون $\mathbf{c}_i$ وابسته خطی است اگر به‌ازای اسکالر ناصفری

چون $v_i \in F$ داشته باشیم $v_i \mathbf{c}_i = \mathbf{0}$؛ با ضرب طرفین در v_i^{-1} (که وجود آن به جهت میدان بودن F حتمی است) به رابطه معادل $\mathbf{c}_i = \mathbf{0}$ می‌رسیم، بنابراین براساس قضیه ۲۷.۷ می‌نیمم فاصله کد برابر $d = 1$ است اگر و فقط اگر $\mathbf{0}$ ستونی از H باشد. دو ستون $\mathbf{c}_i$ و $\mathbf{c}_j$ (با $i \neq j$) وابسته خطی هستند اگر و فقط اگر به‌ازای دو اسکالر v_i و v_j که هر دو صفر نیستند تساوی $v_i \mathbf{c}_i + v_j \mathbf{c}_j = \mathbf{0}$ برقرار باشد. اگر $\mathbf{c}_i$ و $\mathbf{c}_j$ ناصفر باشند آنگاه $v_i v_j \neq 0$. از این‌رو می‌توان این تساوی را به‌فرم $\mathbf{c}_i = a\mathbf{c}_j$ نوشت که $a = -\frac{v_j}{v_i}$. پس دو ستون ناصفر وابسته خطی هستند اگر و فقط اگر هر یک از آنها مضربی از دیگری باشد. (بالاخص، اگر $q = 2$ آنگاه تنها امکان برای a حالت $a = 1$ است، از این‌رو در حالت دوتایی، دو ستون ناصفر وابسته خطی هستند اگر و فقط اگر مساوی باشند.) از قضیه ۲۷.۷ نتیجه می‌شود $d \geq 3$ اگر و فقط اگر ستون‌های H ناصفر بوده و هیچ‌کدام مضربی از دیگری نباشد؛ در حالت دوتایی این شرط معادل با این است که ستون‌های H ناصفر بوده و متمایز باشند.

مثال ۲۸.۷ ماتریس کنترل تساوی $H = (1\ 1\ \dots\ 1)$ برای $\mathcal{P}_n$ دارای ستون‌های ناصفر و مساوی می‌باشد؛ پس می‌نیمم فاصله $\mathcal{P}_n$ برابر $d = 2$ است.

مثال ۲۹.۷ در ماتریس کنترل تساوی

$$H = \begin{pmatrix} 1 & & & -1 \\ & 1 & & -1 \\ & & \ddots & \vdots \\ & & 1 & -1 \end{pmatrix}$$

برای $\mathcal{R}_n$، هر $n-1$ ستونی مستقل خطی هستند، درحالی‌که $\mathbf{c}_1 + \cdots + \mathbf{c}_n = \mathbf{0}$ (متناظر با کدکلمه $11\dots1 = \mathbf{1} \in \mathcal{R}_n$)، از این‌رو $d = n$.

مثال ۳۰.۷ کد دوتایی همینگ $\mathcal{H}_7$ دارای ماتریس کنترل تساوی H است

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

ستون‌ها ناصفر و متمایز هستند، پس $d \geq 3$؛ معادلاً هیچ کدکلمه ناصفر به وزن ۱ یا ۲ وجود ندارد. با این حال، $\mathbf{c}_1 + \mathbf{c}_2 + \mathbf{c}_3 = \mathbf{0}$، پس سه ستون وابسته خطی وجود داشته که متناظر با این واقعیت است که $\mathbf{v} = 1110000$ یک کدکلمه به وزن ۳ است. (این عضو مبنای $\mathbf{u}_1$ برای $\mathcal{H}_7$ در مثال ۵.۶ است.) پس $\mathcal{H}_7$ دارای می‌نیمم فاصله $d = 3$ است.

از ماتریس‌های کنترل تساوی برای ارائه یک اثبات دیگری از کران یکانی برای کدهای خطی (قضیه ۲۳.۷) استفاده می‌شود. اگر H یک ماتریس کنترل تساوی برای یک $[n,k]$-کد

خطی $\mathcal{C}$ باشد، آنگاه $n-k$ سطر آن مستقل خطی هستند؛ از این‌که رتبه سطری و رتبه ستونی در هر ماتریس برابر هستند نتیجه می‌شود H دارای یک مجموعه از $n-k$ ستون مستقل خطی است درحالی‌که هر $n-k+1$ ستونی از H وابسته خطی هستند. درنتیجه براساس قضیه ۲۷.۷ داریم $d \leq n-k+1$.

نتیجه ۳۱.۷ یک $[n,k]$–کد خطی t خطا تصحیح کننده روی F وجود دارد اگر و فقط اگر یک ماتریس H با ابعاد $(n-k) \times n$ و رتبه $n-k$ روی F وجود داشته باشد به‌قسمی‌که هر مجموعه از $2t$ ستون آن مستقل خطی باشند.

اثبات ($\Longrightarrow$) اگر یک چنین کدی مانند $\mathcal{C}$ وجود داشته باشد آنگاه ماتریس کنترل تساوی H برای $\mathcal{C}$ دارای n ستون و $n-k$ سطر مستقل می‌باشد. بنابر قضیه ۱۰.۶، $\mathcal{C}$ دارای می‌نیمم فاصله $d \geq 2t+1$ بوده و بنابر قضیه ۲۷.۷ هر مجموعه از حداکثر $d-1$ ستون H مستقل خطی هستند، پس هر مجموعه از $2t$ ستون مستقل خطی می‌باشند.

($\Longleftarrow$) فرض کنید $\mathcal{V} = F^n$ و $\mathcal{C} = \{\mathbf{v} \in \mathcal{V} \mid \mathbf{v}H^T = \mathbf{0}\}$ و این‌که ماتریس H دارای شرایط لازم باشد؛ پس $\mathcal{C}$ یک کد خطی به‌طول n روی F است. چون رتبه H برابر $n-k$ است سطرهای آن مستقل خطی هستند و از این‌رو بعد $\mathcal{C}$ برابر k است. بنا به فرض، هر مجموعه از ستون‌های مستقل خطی در H حاوی حداقل $2t+1$ ستون بوده و بنابراین قضیه ۲۷.۷ ایجاب می‌کند که $\mathcal{C}$ دارای می‌نیمم فاصله $d \geq 2t+1$ باشد. از این‌رو بنابر قضیه ۱۰.۶ کد $\mathcal{C}$ تعداد t خطا را تصحیح می‌کند. ■

توضیحاتی را در این رابطه در دو بخش بعد به هنگام ساخت کدهای همینگ و گلی[1] ارائه خواهیم کرد.

۴.۷ کدهای همینگ

کد همینگ $\mathcal{H}_7$، $[7,4]$– کد، یک کد دوتایی کامل ۱ خطا تصحیح کننده با نرخ $R = 4/7$ است. در واقع این کد عضوی از یک دنباله نامتناهی از کدهای دوتایی کامل ۱ خطا تصحیح کننده است که در آنها نرخ R با افزایش n به ۱ میل می‌کند. این کدها در ۱۹۵۰ توسط همینگ [Ha 50] معرفی شدند، اگرچه گلی نیز مستقلاً آنها را در همان محدوده زمانی پیدا کرد (برای بحثی پیرامون اولویت در این کشف به [Th 83] مراجعه کنید).

در کران کره‌چینی (نتیجه ۱۷.۶) برای یک کد خطی دوتایی ۱ خطا تصحیح کننده قرار می‌دهیم $t=1$ و $q=2$، و از این‌جا شرط کامل بودن به‌فرم زیر تبدیل می‌شود

$$2^{n-k} = 1 + \binom{n}{1} = 1 + n.$$

با قرار دادن $c = n-k$، این شرط معادل است با

[1] Golay codes

$$n = ۲^c - ۱. \qquad (۴.۷)$$

چون $n - c = ۲^c - ۱ - c$، مقادیر ممکن برای n و k به شرح زیر هستند:

$c =$	۱	۲	۳	۴	۵	...
$n =$	۱	۳	۷	۱۵	۳۱	...
$k =$	۰	۱	۴	۱۱	۲۶	...

حال تلاش می‌کنیم تا کدهایی با این پارامترها بسازیم. برای $t = ۱$، از نتیجه ۳۱.۷ ملاحظه می‌شود که یک چنین کدی وجود دارد اگر و فقط اگر یک ماتریس H $c \times n$ روی $F_۲$ وجود داشته باشد که رتبه آن c بوده و هر دو ستون آن مستقل خطی باشند؛ چون $F = F_۲ = \{۰, ۱\}$، این بدین معنی است که ستون‌های H باید ناصفر و متمایز باشند، از این‌رو H باید متشکل از $n = ۲^c - ۱$ بردار ستونی ناصفر و متمایز به‌طول c باشد. تعداد بردارهای دوتایی ناصفر و متمایز به‌طول c برابر است با $۲^c - ۱$، و بنابراین ستون‌های H دقیقاً تمامی بردارهای ناصفر به‌طول c با ترتیبی دلخواه، می‌باشند (برای حالت $c = ۳$ به مثال ۱۳.۷ مراجعه کنید). این بردارهای ستونی شامل c بردار مبنای استاندارد، که مستقل خطی هستند، نیز می‌باشد و بنابراین رتبه یک ماتریس H این‌چنینی برابر c است. این نشان می‌دهد که $\mathcal{C}$ وجود دارد و هر دو کد خطی با این پارامترها معادل (تحت جایگشت ستونی) هستند. ما این کد $\mathcal{C}$ را کد دوتایی همینگ $\mathcal{H}_n$ به‌طول $n = ۲^c - ۱$ می‌نامیم. به بیان دقیق، ما یک مجموعه از کدهای به‌طول n ساخته‌ایم ولی چون همگی معادل هستند ترجیح داده می‌شود تا به‌صورت غیررسمی آنها را تنها به‌عنوان یک کد $\mathcal{H}_n$ درنظر بگیریم.

مثال ۳۲.۷ از حالت بدیهی $c = ۱$ صرف‌نظر می‌کنیم زیرا $\mathcal{H}_۱$ متشکل است از تنها کدکلمه $\mathbf{0}$ به‌طول ۱. برای $c = ۲$ داریم $n = ۳$ و

$$H = \begin{pmatrix} ۰ & ۱ & ۱ \\ ۱ & ۰ & ۱ \end{pmatrix};$$

پس $\mathcal{H}_۳$ متشکل است از بردارهای دوتایی $\mathbf{v} = v_۱v_۲v_۳$ با خاصیت $v_۲ + v_۳ = v_۱ + v_۳ = ۰$، یا معادلاً $v_۱ = v_۲ = v_۳$؛ پس این کد دوتایی تکراری $\mathcal{R}_۳$ است. برای $c = ۳$ ما کد همینگ $\mathcal{H}_۷$ را داریم که قبلاً ملاحظه شد. برای $c \geq ۴$ تعداد نامتناهی کد کامل $\mathcal{H}_n$ به‌طول $n = ۲^c - ۱$ به‌دست می‌آوریم. نرخ این کدها برابر است با

$$R = \frac{k}{n} = \frac{۲^c - ۱ - c}{۲^c - ۱}$$

که با افزایش c به ۱ میل می‌کند. ولی این کدها تنها یک خطا را تصحیح می‌کنند، و بنابراین $Pr_E \not\longrightarrow ۰$ (تمرین ۴.۷).

تمرین ۴.۷ کد دوتایی همینگ $\mathcal{H}_n$ روی کانال BSC Γ با $P > \frac{1}{2}$ به‌کار می‌رود و Δ کدگشایی نزدیک‌ترین همسایه است. Pr_E را پیدا کنید (تمرین ٦.٦ را برای حالت $n = ۷$ ببینید). وضعیت Pr_E را وقتی n به ∞ میل می‌کند بررسی کنید.

کدگشایی نزدیک ترین همسایه روی $\mathcal{H}_n$ خیلی آسان انجام می‌شود. چون $\mathcal{H}_n$ کامل است و $t = ۱$، پس تمامی خطاهای به وزن حداکثر ۱ تصحیح می‌شوند. فرض کنید $\mathbf{u} \in \mathcal{H}_n$ ارسال و $\mathbf{v} = \mathbf{u} + \mathbf{e}$ دریافت شده است که $wt(\mathbf{e}) \leq ۱$. پس یا $\mathbf{e} = \mathbf{0}$ و یا $\mathbf{e}$ یک بردار مبنای استاندارد از $\mathcal{V}$ است. دریافت کننده $\mathbf{s} = \mathbf{v}H^T$ را که مشخصه[۱] $\mathbf{v}$ نامیده می‌شود محاسبه می‌کند. حال چون $\mathbf{u}H^T = \mathbf{0}$ پس

$$vH^T = (u + e)H^T = uH^T + eH^T = eH^T.$$

بسته به این‌که $\mathbf{e} = \mathbf{0}$ یا $\mathbf{e} \neq \mathbf{0}$ باشد این مشخصه برابر $\mathbf{0}$ یا $\mathbf{c}_i^T$ است. اگر $\mathbf{s} = \mathbf{0}$، دریافت کننده $\mathbf{v}$ را به‌عنوان کلمه ارسال شده درنظر می‌گیرد $(\Delta(\mathbf{v}) = \mathbf{v}(= \mathbf{u})$، و اگر $\mathbf{s} = \mathbf{c}_i^T$ آنگاه $\Delta(\mathbf{v}) = \mathbf{v} - \mathbf{e}_i$ با تغییر سمبل i ام $\mathbf{v}$ به‌دست می‌آید. این روش در حالت $wt(\mathbf{e}) \leq ۱$ همیشه درست عمل می‌کند، ولی اگر $wt(\mathbf{e}) > ۱$ آنگاه عمل کدگشایی هرگز درست نخواهد بود؛ در این حالت $\mathbf{v} = \mathbf{u}' + \mathbf{e}'$ که $\mathbf{u}'$ کدکلمه‌ای متفاوت با $\mathbf{u}$ بوده و $wt(\mathbf{e}') \leq ۱$، و بنابراین الگوریتم یادشده $\mathbf{v}$ را به $\mathbf{u}'$ دکد می‌کند.

مثال ۳۳.۷ ماتریس کنترل تساوی کد $\mathcal{H}_۷$ را که به‌فرم منظم است درنظر بگیرید

$$H = \begin{pmatrix} ۰ & ۱ & ۱ & ۱ & ۱ & ۰ & ۰ \\ ۱ & ۰ & ۱ & ۱ & ۰ & ۱ & ۰ \\ ۱ & ۱ & ۰ & ۱ & ۰ & ۰ & ۱ \end{pmatrix}$$

فرض کنید $\mathbf{u} = ۱۱۰۱۰۰۱ \in \mathcal{H}_۷$ ارسال و $\mathbf{v} = ۱۱۰۱۱۰۱ \in \mathcal{V}$ دریافت شده است، پس $\mathbf{e} = \mathbf{e}_۵$. مشخصه برابر $\mathbf{v}H^T = ۱۰۰$ است که ترانهاده ستون پنجم H است؛ بر این اساس یک خطا در موقعیت پنجم درنظر گرفته شده و بنابراین با تغییر آن در $\mathbf{v}$ نتیجه می‌شود $\Delta(\mathbf{v}) = ۱۱۰۱۰۰۱ = \mathbf{u}$. از طرف دیگر فرض کنید $\mathbf{v}' = ۱۰۰۱۱۰۱$ دریافت شده است و بردار خطا برابر $\mathbf{e}' = \mathbf{e}_۲ + \mathbf{e}_۵$ است. در این‌صورت مشخصه برابر است با $\mathbf{s}' = ۰۰۱ = \mathbf{c}_۷^T$، که اشاره به وجود خطا در موقعیت هفتم دارد، پس $\Delta(\mathbf{v}') = ۱۰۰۱۱۰۰ \neq \mathbf{u}$. بنابراین، به‌جای تصحیح دو خطا عمل کدگشایی یک خطای سومی را نیز ایجاد کرده است.

---

[۱] کلمه Syndrome به معنای مشخصه، واژه‌ای پزشکی است که بر علائمی اطلاق می‌شود که نوع بیماری مریض را مشخص می‌کنند

تمرین ۵.۷ با به‌کار بردن ماتریس کنترل تساوی در مثال ۳۳.۷ نشان دهید $\mathbf{u} = ۱۱۰۰۱۱۰$ یک کدکلمه از $\mathcal{H}_۷$ است. فرض کنید این کلمه ارسال و $\mathbf{v} = ۱۰۰۰۱۱۰$ دریافت شده است. مشخصه را پیدا کرده و $\Delta(\mathbf{v})$ را تعیین کنید. چه اتفاقی خواهد افتاد اگر $\mathbf{v}' = ۰۰۰۰۱۱۰$ دریافت شود، توضیحی در این مورد ارائه دهید.

عمل کدگشایی با $\mathcal{H}_n$ مخصوصاً ساده خواهد بود اگر ستون‌های H به‌گونه‌ای مرتب شده باشند که $\mathbf{c}_۱^T, \ldots, \mathbf{c}_n^T$ به‌ترتیب نمایش اعداد $۱, \ldots, n$ در مبنای ۲ باشند. پس برای $n = ۷$ ماتریس زیر را انتخاب می‌کنیم

$$H = \begin{pmatrix} ۰ & ۰ & ۰ & ۱ & ۱ & ۱ & ۱ \\ ۰ & ۱ & ۱ & ۰ & ۰ & ۱ & ۱ \\ ۱ & ۰ & ۱ & ۰ & ۱ & ۰ & ۱ \end{pmatrix},$$

این ماتریس تحت جایگشت ستونی $(۷\ ۴\ ۵\ ۲\ ۶\ ۳\ ۱)$ معادل ماتریس کنترل تساوی به‌کار رفته در مثال ۳۳.۷ است. یک مشخصه $\mathbf{s} = \mathbf{0}$ به‌عنوان $\mathbf{e} = \mathbf{0}$ تفسیر شده، یعنی این‌که خطایی رخ نداده، در حالی‌که یک مشخصه ناصفر $\mathbf{s}$ نمایش دوتایی مختص i است که خطا در آن‌جا صورت گرفته است.

مثال ۳۴.۷ اجازه دهید نوع معادلی از $\mathcal{H}_۷$ را که توسط ماتریس کنترل تساوی یادشده مشخص می‌شود درنظر بگیریم. با به‌کار بردن جایگشت

$$(۱\ ۳\ ۶\ ۲\ ۵\ ۴\ ۷)$$

روی کلمه ۱۱۰۱۰۰۱ که در مثال ۳۳.۷ به‌کار رفت کدکلمه $\mathbf{u} = ۱۰۱۰۱۰۱ \in \mathcal{H}_۷$ حاصل می‌شود. اگر این کدکلمه ارسال و $\mathbf{v} = ۱۰۱۰۰۰۱$ دریافت شود آنگاه مشخصه برابر $\mathbf{s} = \mathbf{v}H^T = ۱۰۱$ خواهد بود، که این نمایش دوتایی عدد ۵ می‌باشد. از این‌رو با تغییر مؤلفه پنجم $\mathbf{v}$ نتیجه می‌گیریم $\Delta(\mathbf{v}) = ۱۰۱۰۱۰۱ = \mathbf{u}$.

تمرین ۶.۷ مانند مثال ۳۴.۷، نوع معادل $\mathcal{H}_۷$ را به‌کار برده، وضعیت را برای زمانی که $\mathbf{u} = ۰۱۱۱۱۰۰$ ارسال و $\mathbf{v} = ۰۰۱۱۱۰۰$ دریافت شود بررسی کنید.

روش ساخت مشابه‌ای برای کدهای کامل خطی غیردوتایی، $q > ۲$، وجود دارد. در این حالت ستون‌های H را برابر

$$n = \frac{q^c - ۱}{q - ۱} = ۱ + q + q^۲ + \cdots + q^{c-۱} \qquad (۵.۷)$$

بردار به‌طول c روی F_q گرفته به‌قسمی‌که دوبه‌دو نسبت به هم مستقل خطی باشند. (این بیشترین مقدار ممکن برای n می‌باشد؛ تمرین ۷.۷). کد خطی حاصل به‌طول n، بعد $k = n - c$، و می‌نیمم فاصله $d = ۳$ است. از این‌رو $t = ۱$ (تمرین ۷.۷). همچون حالت دوتایی، اگر $c \longrightarrow \infty$ آنگاه $R \longrightarrow ۱$، ولی $Pr_E \not\longrightarrow ۰$.

تمرین ۷.۷ نشان دهید اگر $\mathcal{W} = F_q^c$ آنگاه بیشترین تعداد بردار در $\mathcal{W}$ به‌قسمی‌که دوبه‌دو نسبت به هم مستقل خطی باشند برابر با $(q^c - ۱)/(q - ۱)$ است. نشان دهید اگر یک مجموعه این‌چنینی از بردارها ستون‌های یک ماتریس کنترل تساوی را تشکیل دهند آنگاه کد خطی مربوطه، روی F_q، کامل بوده و ۱ خطا تصحیح کننده است.

مثال ۳۵.۷ اگر $q = ۳$ و $c = ۲$، آنگاه $n = ۴$ و $k = ۲$. می‌توانیم ماتریس

$$H = \begin{pmatrix} ۱ & ۱ & ۱ & ۰ \\ ۱ & ۲ & ۰ & ۱ \end{pmatrix}$$

را درنظر بگیریم که معرف یک $[۴, ۲]$–کد خطی کامل سه‌تایی، روی $F_۳$، بوده و ۱ خطا تصحیح کننده است.

۵.۷ کدهای گلی

گلی با استفاده از نتیجه ۳۱.۷ دو کد کامل $\mathcal{G}_{۱۱}$ و $\mathcal{G}_{۲۳}$ را ساخت که اکنون نام او را با خود دارند. در یک مقاله مشهور [Go 49]، به‌طول تنها نصف صفحه، گلی نه تنها این دو کد بلکه کدهای کامل دوتایی تکراری $\mathcal{R}_n$ (n فرد)، و کدهای کامل ساخته شده در انتهای بخش ۴.۷ را برای تمامی اعداد اول q معرفی کرد (تعمیم به توان اعداد اول مدتی بعد انجام شد).

به‌خاطر بیاورید که یک کد را در صورتی کامل می‌نامند که در کران کره‌چینی خاصیت تساوی برقرار باشد یعنی

$$\sum_{i=۰}^{t} \binom{n}{i} (q-۱)^i = q^{n-k}. \qquad (۶.۷)$$

تمرین ۱۹.۶ اشاره به امکان وجود یک کد کامل با پارامترهای $q = ۳$، $n = ۱۱$، $k = ۶$، و $t = ۲$ دارد. برای ساخت یک چنین کدی گلی ماتریس کنترل تساوی به‌فرم منظم H را درنظر گرفت

$$H = \begin{pmatrix} ۱ & ۱ & ۱ & ۲ & ۲ & ۰ & ۱ & ۰ & ۰ & ۰ & ۰ \\ ۱ & ۱ & ۲ & ۱ & ۰ & ۲ & ۰ & ۱ & ۰ & ۰ & ۰ \\ ۱ & ۲ & ۱ & ۰ & ۱ & ۲ & ۰ & ۰ & ۱ & ۰ & ۰ \\ ۱ & ۲ & ۰ & ۱ & ۲ & ۱ & ۰ & ۰ & ۰ & ۱ & ۰ \\ ۱ & ۰ & ۲ & ۲ & ۱ & ۱ & ۰ & ۰ & ۰ & ۰ & ۱ \end{pmatrix}$$

این ماتریس روی $F_۳$ بوده، با $n = ۱۱$ ستون و $n - k = ۵$ سطر مستقل خطی می‌باشد. با حوصله قابل ملاحظه‌ای می‌توان نشان داد که هیچ چهار ستون (و یا تعداد کمتر) از H

وابسته خطی نبوده درحالی‌که یک مجموعه پنج عضوی از ستون‌های وابسته وجود دارد (به‌عنوان نمونه $\mathbf{c}_2 - \mathbf{c}_7 - \mathbf{c}_8 + \mathbf{c}_9 + \mathbf{c}_{10} = \mathbf{0}$). از قضیه ۲۷.۷ نتیجه می‌شود که کد $\mathcal{C}$، تعریف شده به‌وسیله H، دارای $d = 5$ بوده و در نتیجه $t = 2$. چون

$$\sum_{i=0}^{t} \binom{n}{i}(q-1)^i = 1 + \binom{11}{1}.2 + \binom{11}{2}.2^2 = 243 = 3^5 = q^{n-k},$$

پس این کد کامل است. این کد کامل سه‌تایی $\mathcal{G}_{11}$ است که به‌طول ۱۱ می‌باشد.

مشابهاً با انتخاب $q = 2$، $n = 23$ و $k = 12$ گلی از یک ماتریس کنترل تساوی دوتایی $H = (P^T|I_{11})$ استفاده کرد که

$$P^T = \begin{pmatrix}
1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}.$$

با یک سری عملیات کسل‌کننده می‌توان نشان داد که کمترین تعداد ستون مستقل خطی در H برابر ۷ است، و از این‌رو $d = 7$ و بنابراین $t = 3$. کد متناظر با این ماتریس کنترل تساوی کد گلی دوتایی کامل $\mathcal{G}_{23}$ است که به‌طول $n = 23$ می‌باشد. کامل بودن آن براساس تساوی زیر است:

$$\sum_{i=0}^{t} \binom{n}{i}(q-1)^i = 1 + \binom{23}{1} + \binom{23}{2} + \binom{23}{3} = 2048 = 2^{11} = q^{n-k}.$$

کدهای گلی بسط‌یافته $\mathcal{G}_{12} = \overline{\mathcal{G}}_{11}$ و $\mathcal{G}_{24} = \overline{\mathcal{G}}_{23}$ به‌ترتیب کدهای خطی سه‌تایی [12, 6] و دوتایی [24, 12] می‌باشند. اگرچه این کدها کامل نیستند ولی به لحاظ ارتباط آنها با ساختارهای مهمی در ریاضیات، همچون سیستم‌های اشتاینر[1]، مشبکه[2]، کره‌چینی و گروه‌های ساده [CL 91,CS 92] از جمله کدهای مهم هستند. برای یک منبع جالب در رابطه با پیدایش کدهای گلی و ارتباط آنها با بعضی از این موضوعات به [Th 83] مراجعه کنید. به جهت وجود این ارتباطات راه‌های مختلفی برای ساخت کدهای گلی وجود دارد. اکثر این روش‌ها در مقایسه با روش ساخت اولیه به‌وسیله گلی، که در بالا اشاره شد، دید بیشتری در رابطه با ساختار این کدها می‌دهند اگرچه همگی آنها سرراست نیستند (برای دو مثال نسبتاً ساده، براساس نتایج موجود در [CL 91]، به تمرینات ۱۷.۷ و ۱۸.۷ مراجعه کنید).

[1] Steiner systems [2] Lattices

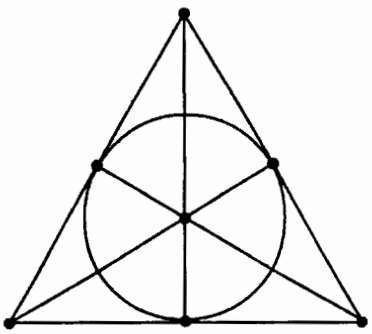
در این‌جا چگونگی به‌دست آوردن کدهای گلی را از سیستم‌های اشتاینر نشان می‌دهیم.
اگر S یک مجموعه n عضوی باشد آنگاه مجموعه توان

$$\mathcal{P}(S) = \{U \mid U \subseteq S\}$$

یک فضای برداری n بعدی روی F_2 است، که در آن مجموع دو زیرمجموعه U و V برابر تفریق متقارن آنها می‌باشد، یعنی این‌که $U + V := (U \cup V) \setminus (U \cap V)$، و عضو صفر برابر مجموعه خالی $\emptyset$ است. اگر $S = \{s_1, \ldots, s_n\}$ آنگاه هر زیرمجموعه U قابل نمایش به‌صورت یک بردار $\mathbf{u} = u_1 \ldots u_n \in \mathcal{V} = F_2^n$ است که در آن $u_i = 1$ اگر و فقط اگر $s_i \in U$. بنابراین می‌توان هر زیرمجموعه غیرخالی $\mathcal{C} \subseteq \mathcal{P}(S)$ را به‌عنوان یک کد دوتایی به‌طول n درنظر گرفت؛ این کد خطی است اگر و فقط اگر تحت عمل جمع بسته باشد. با ملاحظه $wt(\mathbf{u}) = |U|$ و $d(\mathbf{u}, \mathbf{v}) = |U + V|$ می‌بینیم برای این‌که می‌نیمم فاصله بزرگ باشد باید $\mathcal{C}$ به‌گونه‌ای انتخاب شود که هر دو زیرمجموعه متمایز $U, V \in \mathcal{C}$ به‌قدر کافی با هم متفاوت باشند.

یک روش سیستماتیک برای انجام این کار استفاده از طرح‌های قالبی است. یک t طرح روی یک مجموعه S یک مجموعه از زیرمجموعه‌های S، به نام قالب، است که همگی دارای اندازه یکسان بوده و به‌ازای عددی چون λ هر زیرمجموعه t عضوی از S دقیقاً زیرمجموعه λ قالب باشد. این شرایط محدودیت‌های قوی روی کد حاصل ایجاد می‌کنند. روابط بین طرح‌ها و کدها به‌تفصیل در [CL 91] توضیح داده شده و در این‌جا توجه خود را تنها به چند مثال ساده محدود می‌کنیم. با نوشتن l به‌جای t (که ما قبلاً آن‌را به‌عنوان تعداد خطاهای تصحیح شده به‌کار برده‌ایم) یک سیستم اشتاینر را به‌عنوان یک l–طرح با $\lambda = 1$ تعریف می‌کنیم؛ یعنی این‌که یک سیستم اشتاینر مجموعه‌ای از قالب‌های m عضوی B از یک مجموعه n عضوی S است به‌قسمی‌که هر مجموعه l عضوی از S دقیقاً در یک قالب قرار دارد. یک چنین سیستمی را با $S(l, m, n)$ نمایش می‌دهیم.

مثال ۳۶.۷ فرض کنید S مجموعه تمامی زیرفضاهای یک بعدی فضای برداری $\mathcal{W} = F^c$ باشد که $c \geq 2$. پس $|S| = (q^c - 1)/(q - 1)$. هر زیرفضای دو بعدی از $\mathcal{W}$ حاوی $q + 1$ عضو S است. این زیرمجموعه‌های $q + 1$ عضوی را به‌عنوان قالب درنظر می‌گیریم. هر زوج متمایز از زیرفضاهای یک بعدی از $\mathcal{W}$ یک زیرفضای دو بعدی منحصربه‌فرد را تولید می‌کنند. از این‌رو هر زوج متمایز از اعضای S دقیقاً در یک قالب قرار دارند. بنابراین یک سیستم اشتاینر داریم که در آن $l = 2$، $m = q + 1$ و $n = (q^c - 1)/(q - 1)$؛ این هندسه تصویری $PG(c - 1, q)$ است که خطوط این هندسه همان قالب‌ها هستند. در شکل ۱.۷ هفت قالب سه عضوی صفحه فانو $PG(2, 2)$ را ملاحظه می‌کنید؛ برای روابط بین این هندسه‌ها و کدهای همینگ به تمرین ۱۲.۷ مراجعه کنید.



شکل ۱.۷

اگر یک سیستم اشتاینر $S(l,m,n)$ دارای b قالب باشد آنگاه

$$\binom{n}{l} = b\binom{m}{l}.$$

علت این تساوی این است که تعداد زیرمجموعه‌های l عضوی S برابر $\binom{n}{l}$ بوده و هر یک از آنها در یک قالب منحصربه‌فرد قرار گرفته، و هر قالب حاوی $\binom{m}{l}$ زیرمجموعه این‌چنینی است. پس $\binom{m}{l}$ یک بخش‌کننده $\binom{n}{l}$ بوده و از این‌رو یک محدودیت روی پارامترهای m، n و l ایجاد می‌شود. در واقع محدودیت‌های بیشتری وجود دارند. اگر $s \in S$، آنگاه به‌سادگی می‌توان دید که $S' = S\backslash\{s\}$ یک سیستم اشتاینر $S(l-1, m-1, n-1)$ است که در آن قالب‌ها مجموعه‌های $B\backslash\{s\}$ بوده که B یک قالب شامل s در سیستم اشتاینر $S(l,m,n)$ است. در این‌صورت از بحث قبلی نتیجه می‌شود که $\binom{m-1}{l-1}$ یک بخش‌کننده $\binom{n-1}{l-1}$ است. با ادامه این بحث محدودیت‌های بیشتری حاصل می‌شود.

مثال ۳۷.۷ اگر S یک سیستم اشتاینر $S(2,3,n)$ باشد، آنگاه بر پایه این دو شرط ۳ بخش‌کننده $n(n-1)/2$ و ۲ بخش‌کننده $n-1$ است، پس n به هنگ ۶ برابر ۱ یا ۳ است. در واقع این شرط لازم برای وجود S یک شرط کافی نیز می‌باشد [قضیه ۳.۴.۱۵ در Ha 67]؛ هندسه $PG(c-1,2)$ مثالی برای $n = 2^c - 1$ است.

مثال ۳۸.۷ سه‌تایی $(5,8,24)$ در شرط لازم یادشده صدق می‌کند، بنابراین وجود یک سیستم اشتاینر $S(5,8,24)$ با $b = \binom{24}{5}/\binom{8}{5} = 759$ قالب موجه به‌نظر می‌رسد. وجود یک چنین سیستمی و یکتایی آن ثابت شده است؛ گروه اتومورفیسم آن (مجموعه‌ای از جایگشت‌های S که قالب را به قالب می‌برند) گروه M_{24}[1] است که یک گروه ساده از مرتبه 244823040 است. حال فرض کنید $\mathcal{C}$ زیرفضای تولید شده توسط قالب‌های S در فضای $\mathcal{V} = \mathcal{P}(S) = F_2^{24}$ باشد. با درنظر گرفتن تعریف سیستم اشتاینر، بدون نیاز به دانستن

[1] Mathieu group M_{24}

قالب‌ها، می‌توان با به‌کار بردن روش‌های ساده شمارش نشان داد که $\mathcal{C}$ متشکل است از:
۱ مجموعه با اندازه ۰ که همان $\emptyset$ می‌باشد؛
۷۵۹ مجموعه B به اندازه ۸، که قالب‌ها هستند؛
۲۵۷۶ مجموعه $B + B'$ به اندازه ۱۲، که B و B' قالب‌های با خاصیت $|B \cap B'| = ۲$ هستند؛
۷۵۹ مجموعه $B + B'$ به اندازه ۱۶، که B و B' قالب‌های جدا از هم هستند؛
۱ مجموعه با اندازه ۲۴، که S بوده و مجموع سه قالب جدا از هم می‌باشد.
(برای جزییات بیشتر بخش ۳.۷ از [An 74] را ملاحظه کنید.) چون

$$۱ + ۷۵۹ + ۲۵۷۶ + ۷۵۹ + ۱ = ۴۰۹۶ = ۲^{۱۲}$$

پس $\mathcal{C}$ یک $[۲۴, ۱۲]$–کد خطی است. این در واقع کد گلی بسط یافته $\mathcal{G}_{۲۴}$ است. کدکلمه‌ها به وزن ۰، ۸، ۱۲، ۱۶ و ۲۴ هستند و بنابراین $d = ۸$. با پنچر کردن $\mathcal{G}_{۲۴}$ در مختص دلخواه i (حذف مختص i ام در تمامی کدکلمه‌ها) یک $[۲۳, ۱۲]$–کد خطی با $d = ۷$ به‌دست می‌آید؛ این کد حاصل کد گلی کامل $\mathcal{G}_{۲۳}$ است؛ انتخاب مختص i اهمیتی ندارد زیرا تمامی کدهای به‌دست آمده معادل هستند.

تمرین ۸.۷ ثابت کنید در یک سیستم اشتاینر $S = S(۵, ۸, ۲۴)$ هر عضو $s \in S$ در ۲۵۳ قالب قرار می‌گیرد، هر دو عضوی در ۷۷ قالب، هر سه عضوی در ۲۱ قالب، و سرانجام هر چهار عضوی در ۵ قالب قرار می‌گیرد.

می‌توان جهت بحث را عوض کرده و سیستم اشتاینر را از روی کد به‌دست آورد: قالب‌های $S(۵, ۸, ۲۴)$ محمل‌های $U = \{i \mid u_i \neq ۰\}$ از کدکلمات به وزن ۸ در $\mathcal{G}_{۲۴}$ هستند (برای این نوع نگاه به مسئله به [CL 91] مراجعه کنید). مشابهاً محمل کلمات به وزن ۷ در $\mathcal{G}_{۲۳}$ تشکیل یک سیستم اشتاینر $S(۴, ۷, ۲۳)$ می‌دهند، درحالی‌که کلمات به وزن ۵ در $\mathcal{G}_{۱۱}$ و به وزن ۶ در $\mathcal{G}_{۱۲}$ به‌ترتیب تشکیل سیستم‌های اشتاینر $S(۴, ۵, ۱۱)$ و $S(۵, ۶, ۱۲)$ می‌دهند. با این حال، برای این دو حالت آخر استخراج کد از روی طرح پیچیده‌تر است که در واقع این وضعیت برای اکثر حالت‌های غیردوتایی صادق است.

۶.۷ آرایش استاندارد

در کدگشایی نزدیک‌ترین همسایه، برای یک بردار داده شده $\mathbf{v} \in \mathcal{V}$ نیاز به توانایی پیدا کردن کدکلمه $\mathbf{u} = \Delta(\mathbf{v}) \in \mathcal{C}$ داریم که کمترین فاصله را با $\mathbf{v}$ داشته باشد. وقتی $\mathcal{C}$ یک کد خطی است یک الگوریتم بر پایه آرایش استاندارد[۱] برای انجام این کار موجود است. در واقع در این

[۱] Standard array

الگوریتم کلمات فضا در یک جدول متشکل از هم‌دسته‌های زیرفضای $\mathcal{C}$ نوشته می‌شوند. فرض کنید

$$\mathcal{C} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M\}$$

یک کد خطی با $M = q^k$ عضو باشد؛ کلمه $\mathbf{0}$ باید یک کدکلمه باشد، از این‌رو شماره‌گذاری را به‌گونه‌ای انتخاب می‌کنیم که $\mathbf{u}_1 = \mathbf{0}$. برای $i \geq 1$، سطر i ام آرایش استاندارد را با انتخاب کلمه $\mathbf{v}_i$ که کمترین وزن را داشته و در سطرهای قبل ظاهر نشده شروع می‌کنیم (بالاخص $\mathbf{v}_1 = \mathbf{0}$)؛ سپس سطر i ام را برابر هم‌دسته $\mathbf{v}_i + \mathcal{C}$ قرار می‌دهیم

$$\mathbf{v}_i + \mathcal{C} = \{\mathbf{v}_i + \mathbf{u}_1 (= \mathbf{v}_i),\ \ \mathbf{v}_i + \mathbf{u}_2, \ldots, \mathbf{v}_i + \mathbf{u}_M\}$$

که عناصر آن به‌ترتیب بالا نوشته می‌شوند. پس اولین سطر برابر $\mathbf{v}_1 + \mathcal{C} = \mathbf{0} + \mathcal{C} = \mathcal{C}$ است، سطرهای متمایز از هم جدا هستند، و عملیات پس از تشکیل $q^n/M = q^{n-k}$ سطر متوقف می‌شود. پس از اتمام عملیات، هر $\mathbf{v} \in \mathcal{V}$ دقیقاً یک‌بار در آرایش به‌صورت

$$\mathbf{v} = \mathbf{v}_i + \mathbf{u}_j, \tag{۷.۷}$$

ظاهر می‌شود، به‌قسمی‌که $\mathbf{v}$ برابر j امین عنصر سطر i ام می‌باشد. اعضای $\mathbf{v}_i$ نمایندگان هم‌دسته‌های $\mathcal{C}$ در $\mathcal{V}$ می‌باشند که سردسته نامیده می‌شوند. براساس روش ساخت آرایش استاندارد داریم

$$wt(\mathbf{v}_1) \leq wt(\mathbf{v}_2) \leq wt(\mathbf{v}_3) \leq \ldots;$$

دقیقاً پس از آخرین سطر با خاصیت $wt(\mathbf{v}_i) \leq t$، $t = \lfloor \frac{d-1}{2} \rfloor$، یک خط افقی در طول آرایش رسم می‌کنیم. توجه کنید که آرایش استاندارد منحصربه‌فرد نیست، در سطر i ام ممکن است بیش از یک کلمه با وزن می‌نیمم وجود داشته باشد و بنابراین سردسته $\mathbf{v}_i$ منحصربه‌فرد نباشد.

مثال ۳۹.۷ فرض کنید $\mathcal{C}$ کد خطی دوتایی تکراری $\mathcal{R}_4$ به‌طول $n = 4$ باشد، پس $q = 2$، $k = 1$ و کدکلمات عبارت هستند از $\mathbf{u}_1 = \mathbf{0} = 0000$ و $\mathbf{u}_2 = \mathbf{1} = 1111$. تعداد $q^{n-k} = 8$ هم‌دسته وجود دارد که هر یک شامل ۲ عضو می‌باشد؛ پس آرایش استاندارد دارای ۸ سطر و ۲ ستون است. مجبور به انتخاب $\mathbf{v}_1 = \mathbf{0}$ به‌عنوان اولین سردسته هستیم؛ چهار سردسته بعدی اعضای مبنای استاندارد هستند که به وزن ۱ می‌باشند، و سه سردسته آخر (که منحصربه‌فرد نیستند) به وزن ۲ هستند. این کد دارای $d = 4$ و بنابراین $t = 1$ است، پس خط افقی را زیر سطر پنجم رسم می‌کنیم. به‌عنوان نمونه، یک شکل ممکن برای

آرایش استاندارد به‌فرم زیر است:

۰۰۰۰	۱۱۱۱
۱۰۰۰	۰۱۱۱
۰۱۰۰	۱۰۱۱
۰۰۱۰	۱۱۰۱
۰۰۰۱	۱۱۱۰
-----	-----
۱۱۰۰	۰۰۱۱
۱۰۱۰	۰۱۰۱
۱۰۰۱	۰۱۱۰

لم ۴۰.۷

(الف) اگر $\mathbf{v}$ در ستون j ام آرایش استاندارد قرار داشته باشد (یعنی این‌که به‌ازای عددی چون i داشته باشیم $\mathbf{v} = \mathbf{v}_i + \mathbf{u}_j$)، آنگاه $\mathbf{u}_j$ نزدیک‌ترین کلمه به $\mathbf{v}$ است.

(ب) افزون بر این، اگر $\mathbf{v}$ بالای خط رسم شده در آرایش استاندارد باشد (یعنی این‌که $wt(\mathbf{v}_i) \leq t$)، آنگاه $\mathbf{u}_i$ کدکلمه منحصربه‌فرد نزدیک به $\mathbf{v}$ است.

اثبات

(الف) فرض کنید $\mathbf{v} = \mathbf{v}_i + \mathbf{u}_j$، و فرض کنید که $\mathbf{u}_j$ نزدیک‌ترین کدکلمه به $\mathbf{v}$ نباشد، پس کدکلمه دیگری چون $\mathbf{u}_{j'}$ موجود است که $d(\mathbf{v}, \mathbf{u}_{j'}) < d(\mathbf{v}, \mathbf{u}_j)$. چون $d(\mathbf{v}, \mathbf{u}) = wt(\mathbf{v} - \mathbf{u})$ و $\mathbf{v} - \mathbf{u}_j = \mathbf{v}_i$ پس

$$wt(\mathbf{v} - \mathbf{u}_{j'}) < wt(\mathbf{v} - \mathbf{u}_j) = wt(\mathbf{v}_i);$$

حال از $\mathbf{u}_j - \mathbf{u}_{j'} \in \mathcal{C}$ نتیجه می‌شود

$$\mathbf{v} - \mathbf{u}_{j'} = \mathbf{v}_i + \mathbf{u}_j - \mathbf{u}_{j'} \in \mathbf{v}_i + \mathcal{C}$$

که در تناقض با این است که $\mathbf{v}_i$ در هم‌دسته مربوط به خود کمترین وزن را دارد.

(ب) علاوه بر شرایط فوق فرض کنید $wt(\mathbf{v}_i) \leq t$ و فرض کنید به‌ازای عضوی چون $\mathbf{u}_{j'} \in \mathcal{C}$ داشته باشیم $d(\mathbf{v}, \mathbf{u}_{j'}) \leq d(\mathbf{v}, \mathbf{u}_j)$. در این‌صورت

$$\begin{aligned}
d(\mathbf{u}_j, \mathbf{u}_{j'}) &\geq d && \text{(بنابه تعریف } d\text{)} \\
&\geq 2t && \text{(بنابر قضیه ۱۰.۶)} \\
&\geq 2d(\mathbf{v}, \mathbf{u}_j) && (wt(\mathbf{v}_i) \leq t \text{ چون}) \\
&\geq d(\mathbf{v}, \mathbf{u}_j) + d(\mathbf{v}, \mathbf{u}_{j'}) && (d(\mathbf{v}, \mathbf{u}_{j'}) \leq d(\mathbf{v}, \mathbf{u}_j) \text{ چون}) \\
&\geq d(\mathbf{u}_j, \mathbf{u}_{j'}), && \text{(بنابر نامساوی مثلث)}
\end{aligned}$$

پس $d(\mathbf{u}_j, \mathbf{u}_{j'}) > d(\mathbf{u}_j, \mathbf{u}_{j'})$ که یک تناقض است. ■

براین اساس کره $S_t(\mathbf{u}_j)$، تعریف شده در بخش ۴.۶، دقیقاً برابر قسمتی از ستون j ام است که بالای خط افقی قرار می‌گیرد. بنابراین $\mathcal{C}$ کامل است اگر و فقط اگر تمامی آرایش استاندارد بالای خط افقی باشد.

می‌توان از لم ۴۰.۷ برای کدگشایی استفاده کرد. فرض کنید کدکلمه $\mathbf{u} \in \mathcal{C}$ ارسال و $\mathbf{v} = \mathbf{u} + \mathbf{e} \in \mathcal{V}$ دریافت شده است که $\mathbf{e}$ کلمه خطا می‌باشد. گیرنده $\mathbf{v} = \mathbf{v}_i + \mathbf{u}_j$ را در آرایش استاندارد پیدا کرده و $\Delta(\mathbf{v}) = \mathbf{u}_j$ را به‌عنوان محتمل‌ترین کدکلمه ارسال شده درنظر می‌گیرد زیرا این یکی از نزدیک‌ترین کدکلمه‌ها به $\mathbf{v}$ است (در واقع اگر $\mathbf{v}$ بالای خط باشد $\mathbf{u}_j$ نزدیک‌ترین کدکلمه به $\mathbf{v}$ است). پس هر کلمه دریافتی به کدکلمه $\mathbf{u}_j$ که در همان ستون مربوط به $\mathbf{v}$ قرار دارد دکد می‌شود. این تصمیم درست است اگر و فقط اگر $\mathbf{u} = \mathbf{u}_j$، یعنی این‌که اگر و فقط اگر $\mathbf{e} = \mathbf{v}_i$. پس این قاعده منجر به کدگشایی درست می‌شود اگر و فقط اگر کلمه خطا یک سردسته باشد.

مثال ۴۱.۷ فرض کنید $\mathcal{C} = \mathcal{R}_4$. آرایش استاندارد مثال ۳۹.۷ را درنظر بگیرید. نیز فرض کنید $\mathbf{u} = ۱۱۱۱$ ارسال شده و کلمه خطا برابر $\mathbf{e} = ۰۱۰۰$ باشد (پس تنها سمبل دوم کلمه $\mathbf{u}$ نادرست ارسال شده است). در این‌صورت $\mathbf{v} = ۱۱۱۱ + ۰۱۰۰ = ۱۰۱۱$ دریافت شده و چون $\mathbf{v}$ در ستون مربوط به $\mathbf{u} = ۱۱۱۱$ قرار دارد، گیرنده کلمه ارسال شده را برابر $\Delta(\mathbf{v}) = ۱۱۱۱$ درنظر می‌گیرد. با این حال، اگر کلمه خطا برابر $\mathbf{e} = ۰۱۱۰$ باشد آنگاه کلمه $\mathbf{v} = ۱۰۰۱$ دریافت می‌شود؛ این در ستون مربوط به $\mathbf{u}_1 = ۰۰۰۰$ قرار دارد، پس گیرنده به غلط تصمیم به $\Delta(۱۰۰۱) = ۰۰۰۰$ می‌گیرد. در واقع این آرایش استاندارد تمامی خطاهای به وزن حداکثر ۱ را تصحیح کرده ولی از میان خطاهای به وزن ۲ تنها خطاهای ۱۱۰۰، ۱۰۱۰ و ۱۰۰۱ را تصحیح می‌کند درحالی‌که هیچ خطای به وزن ۳ یا ۴ را تصحیح نمی‌کند. هر انتخابی از آرایش استاندارد سه کلمه خطای به وزن ۲ را تصحیح می‌کند ولی نه لزوماً این سه کلمه را.

نقطه قوت این روش کدگشایی سادگی نسبی در فهم و پیاده‌سازی آن است. نقطه ضعف آن نیاز به حافظه بالا (آرایش استاندارد حاوی تمامی کلمات فضا است) و زمان‌بر بودن عمل پیدا کردن کلمه دریافتی در آرایش می‌باشد. از این‌رو در بخش بعد یک روش کدگشایی معادل ولی کاراتر برای کدهای خطی ارائه می‌دهیم.

۷.۷ کدگشایی مشخصه

کدگشایی مشخصه[۱] یک نوع طبیعی‌تر از روش آرایش استاندارد است. اگر H یک ماتریس کنترل تساوی برای یک کد خطی $\mathcal{C} \subseteq \mathcal{V}$ باشد، آنگاه مشخصه یک کلمه دریافتی $\mathbf{v} \in \mathcal{V}$

[۱] Syndrom decoding

برابر است با

$$\mathbf{s} = \mathbf{v}H^T \in F^{n-k}; \qquad (۸.۷)$$

(از این ایده در بخش ۴.۷ در رابطه با کدهای دوتایی همینگ استفاده کردیم). پس $\mathbf{s} = s_1 \dots s_{n-k}$، که $s_i = \mathbf{v}.\mathbf{r}_i$ و $\mathbf{r}_i$ سطر i ام ماتریس H است؛ بنابراین s_i نتیجه اعمال i امین معادله کنترل تساوی روی $\mathbf{v}$ است. نتیجه بعد نشان می‌دهد که مشخصه $\mathbf{s}$ هم‌دسته حاوی $\mathbf{v}$ را تعیین می‌کند، یا معادلاً سطر حاوی $\mathbf{v}$ در آرایش استاندارد را مشخص می‌کند. به‌خاطر بیاورید که بردارهای $\mathbf{v}$ و $\mathbf{v}'$ در یک هم‌دسته از زیرفضای $\mathcal{C}$ قرار می‌گیرند (یعنی $\mathbf{v} + \mathcal{C} = \mathbf{v}' + \mathcal{C}$) اگر و فقط اگر $\mathbf{v} - \mathbf{v}' \in \mathcal{C}$.

لم ۴۲.۷ فرض کنید $\mathcal{C}$ یک کد خطی با ماتریس کنترل تساوی H بوده، و فرض کنید کلمات $\mathbf{v}$ و $\mathbf{v}'$ از $\mathcal{V}$ دارای مشخصه‌های $\mathbf{s}$ و $\mathbf{s}'$ باشند. در این‌صورت $\mathbf{v}$ و $\mathbf{v}'$ در یک هم‌دسته از $\mathcal{C}$ قرار می‌گیرند اگر و فقط اگر $\mathbf{s} = \mathbf{s}'$.

اثبات

$$\begin{aligned}
\mathbf{v} + \mathcal{C} = \mathbf{v}' + \mathcal{C} &\iff \mathbf{v} - \mathbf{v}' \in \mathcal{C} \\
&\iff (\mathbf{v} - \mathbf{v}')H^T = \mathbf{0} \qquad (\text{بنابر لم } ۱۰.۷) \\
&\iff \mathbf{v}H^T = \mathbf{v}'H^T \\
&\iff \mathbf{s} = \mathbf{s}'. \ \blacksquare
\end{aligned}$$

این نشان می‌دهد که یک بردار $\mathbf{v} \in \mathcal{V}$ در i امین سطر آرایش استاندارد قرار می‌گیرد اگر و فقط اگر مشخصه آن با مشخصه $\mathbf{v}_i$ یکی باشد، یعنی $\mathbf{v}H^T = \mathbf{v}_iH^T$. از این‌رو یک جدول مشخصه متشکل از دو ستون ایجاد می‌کنیم؛ سردسته‌های $\mathbf{v}_i$ (که در بخش ۶.۷ انتخاب شدند) در اولین ستون و مشخصه‌های آنها $\mathbf{s}_i = \mathbf{v}_iH^T$ در ستون دوم قرار می‌گیرند.

مثال ۴۳.۷ فرض کنید $\mathcal{C}$ کد تکراری $\mathcal{R}_۴$ با آرایش استاندارد ارائه شده در مثال ۳۹.۷ باشد. پس سردسته‌ها کلمات ستون اول هستند. با به‌کار بردن ماتریس کنترل تساوی H که در مثال ۱۱.۷ معرفی شده است

$$H = \begin{pmatrix} ۱ & & & -۱ \\ & ۱ & & -۱ \\ & & ۱ & -۱ \end{pmatrix} = \begin{pmatrix} ۱ & & & ۱ \\ & ۱ & & ۱ \\ & & ۱ & ۱ \end{pmatrix}$$

و اعمال آن روی هر بردار $\mathbf{v} = v_1v_2v_3v_4 \in \mathcal{V}$ ملاحظه خواهیم کرد که $\mathbf{s} = \mathbf{v}H^T = s_1s_2s_3$ که $s_i = v_i + v_4$ برای $1 \le i \le 3$. با اعمال این روی سردسته‌های $\mathbf{v}_1$، ... ، $\mathbf{v}_8$ مشخصه‌های متناظر با آنها به‌دست می‌آید. از این‌جا جدول مشخصه زیر

حاصل می‌شود:

$\mathbf{v}_i$	$\mathbf{s}_i$
۰۰۰۰	۰۰۰
۱۰۰۰	۱۰۰
۰۱۰۰	۰۱۰
۰۰۱۰	۰۰۱
۰۰۰۱	۱۱۱
۱۱۰۰	۱۱۰
۱۰۱۰	۱۰۱
۱۰۰۱	۰۱۱

در حالت کلی، اگر یک ماتریس کنترل تساوی H و یک جدول مشخصه برای یک کد خطی $\mathcal{C}$ در دست باشد، آنگاه عمل کدگشایی به شرح زیر انجام می‌شود. برای یک بردار دریافت شده $\mathbf{v}$ مشخصه آن، $\mathbf{s} = \mathbf{v}H^T$، را محاسبه کرده و سپس $\mathbf{s}$ را در ستون دوم جدول مشخصه پیدا می‌کنیم؛ فرض کنید $\mathbf{s} = \mathbf{s}_i$. اگر $\mathbf{v}_i$ سردسته متناظر با $\mathbf{s}_i$ باشد آنگاه بنابر لم ۴۲.۷ بردارهای $\mathbf{v}$ و $\mathbf{v}_i$ در یک هم‌دسته قرار دارند، از این‌رو به‌ازای کدکلمه‌ای چون $\mathbf{u}_j$ داریم $\mathbf{v} = \mathbf{v}_i + \mathbf{u}_j$. بنابراین کلمه $\mathbf{v}$ را به‌عنوان $\mathbf{u}_j$ دکد می‌کنیم. پس $\Delta(\mathbf{v}) = \mathbf{u}_j = \mathbf{v} - \mathbf{v}_i$ که $\mathbf{v}H^T = \mathbf{s}_i$.

مثال ۴۴.۷ مجدداً فرض کنید $\mathcal{C}$ کد تکراری $\mathcal{R}_۴$ با ماتریس کنترل تساوی H و جدول مشخصه ارائه شده در مثال ۴۳.۷ باشد. اگر $\mathbf{v} = ۱۱۰۱$ دریافت شود ابتدا مشخصه $\mathbf{s} = \mathbf{v}H^T = ۰۰۱$ را پیدا می‌کنیم. این برابر $\mathbf{s}_۴$ در جدول مشخصه است، پس

$$\Delta(\mathbf{v}) = \mathbf{v} - \mathbf{v}_۴ = ۱۱۰۱ - ۰۰۱۰ = ۱۱۱۱.$$

نقطه قوت این روش در این است که با معلوم بودن H و جدول مشخصه عمل کدگشایی نسبتاً سریع است؛ با داشتن $\mathbf{v}$ مشخصه $\mathbf{s} = \mathbf{v}H^T$ به‌سادگی محاسبه می‌شود؛ چون جدول مشخصه خیلی کوچک‌تر از آرایش استاندارد است پیدا کردن $\mathbf{s}$ در این جدول به مراتب آسان‌تر از پیدا کردن $\mathbf{v}$ در آرایش استاندارد است، به‌ویژه اگر مشخصه‌ها با ترتیب خاصی در جدول مرتب شده باشند؛ سرانجام این‌که محاسبه $\mathbf{u}_j = \mathbf{v} - \mathbf{v}_i$ آسان است.

مثال ۴۵.۷ می‌توان الگوریتم کدگشایی کدهای همینگ، بخش ۴.۷، را به زبان جدول مشخصه بیان کرد. اگر $\mathcal{C}$ یک کد دوتایی همینگ $\mathcal{H}_n$ باشد، آنگاه سردسته‌های $\mathbf{v}_i$ برابر $n+۱$ بردار $\mathbf{v} \in \mathcal{C}$ به وزن $wt(\mathbf{v}) \leq ۱$ هستند که با $\mathbf{v}_۱ = \mathbf{0}$ شروع شده و با ترتیب خاصی از اعضای مبنای استاندارد نوشته می‌شوند. مشخصه‌های $\mathbf{s}_i$ با $\mathbf{s}_۱ = \mathbf{0}$ شروع شده و با ترانهاده ستون‌های H تحت ترتیب یادشده تکمیل می‌شود. اجازه دهید این ستون‌ها به‌ترتیب نمایش دوتایی اعداد $۱، ۲، \ldots، n$ بوده و سردسته‌ها به‌ترتیب $\mathbf{v}_{i+۱} = \mathbf{e}_i$، $۱ \leq i \leq n$، باشند؛

در این‌صورت مشخصه‌های $s_1, \ldots, s_{n+1}$ به‌ترتیب نمایش دوتایی اعداد $0, 1, \ldots, n$ هستند. اگر یک کلمه دریافتی $\mathbf{v}$ مشخصه $\mathbf{s} = \mathbf{0}$ را تولید کند این بدین معنی خواهد بود که هیچ خطایی صورت نگرفته و بنابراین $\mathbf{v} = \Delta(\mathbf{v})$؛ از طرف دیگر، یک مشخصه $\mathbf{s} \neq \mathbf{0}$ به‌عنوان نمایش دوتایی موقعیتی چون i که در آن یک خطا صورت گرفته است لحاظ می‌شود، پس $\Delta(\mathbf{v}) = \mathbf{v} - \mathbf{e}_i$.

تمرین ۹.۷ فرض کنید $\mathcal{C}$ کد خطی دوتایی تولید شده توسط بردارهای ۰۱۱۰۱۱، ۱۰۱۱۰۱ و ۱۱۱۰۰۰ باشد. یک ماتریس مولد منظم برای $\mathcal{C}$ نوشته و از آن‌جا یک ماتریس کنترل تساوی H برای $\mathcal{C}$ بنویسید. کدکلمه $\mathbf{c}$ متناظر با ارقام اطلاعات ۱۱۰ را پیدا کنید و درستی $\mathbf{c}H^T = \mathbf{0}$ را بررسی کنید. نرخ R و می‌نیمم فاصله d را برای $\mathcal{C}$ پیدا کنید. یک جدول مشخصه برای $\mathcal{C}$ تعیین کنید؛ چه خطاهایی را این جدول تصحیح می‌کند؟ مقدار Pr_E را پیدا کنید که در آن کانال Γ یک BSC با $P > \frac{1}{2}$ است.

۸.۷ تمرینات تکمیلی

تمرین ۱۰.۷ نشان دهید که تعداد کدهای خطی متمایز k–بعدی $\mathcal{C} \subseteq \mathcal{V} = F_q^n$ برابر است با

$$\frac{(q^n - 1)(q^n - q) \ldots (q^n - q^{n-k+1})}{(q^k - 1)(q^k - q) \ldots (q^k - q^{k-1})}.$$

تمرین ۱۱.۷ نشان دهید اگر L_1 و L_2 دو خط متمایز در صفحه فانو $S = PG(2, 2)$ باشند آنگاه تفریق متقارن آنها $L_1 + L_2$ مکمل خط سوم دیگری است. نتیجه بگیرید زیرفضای $\mathcal{C}$ از $\mathcal{V} = \mathcal{P}(S) \cong F_2^7$ که توسط خط‌ها تولید می‌شود متشکل است از $\emptyset$، ۷ خط، مکمل این ۷ خط، و S. نشان دهید که این کد معادل همینگ کد $\mathcal{H}_7$ است.

تمرین ۱۲.۷ نشان دهید اگر $\mathcal{C}$ یک کد دوتایی کامل ۱ خطا تصحیح کننده به‌طول n باشد آنگاه محمل کدکلمات به وزن ۳ معرف قالب‌های یک سیستم اشتاینر $S(2, 3, n)$ روی $S = \{1, 2, \ldots, n\}$ هستند. نشان دهید اگر $\mathcal{C} = \mathcal{H}_n$، که $n = 2^c - 1$، آنگاه سیستم اشتاینر حاصل ایزومورف با PG $(c - 1, 2)$ است.

تمرین ۱۳.۷ یک اتومورفیسم از یک کد $\mathcal{C} \subseteq \mathcal{V} = F^n$ یک جایگشت روی مختصات است که تحت آن $\mathcal{C}$ به خودش نگاشته می‌شود. نشان دهید که اتومورفیسم‌های $\mathcal{C}$، Aut $(\mathcal{C})$، یک زیرگروه از گروه متقارن S_n است. گروه‌های Aut $(\mathcal{R}_n)$ و Aut $(\mathcal{P}_n)$ را تعیین کنید. کلمات کد دوتایی $\mathcal{R}_2 \oplus \mathcal{R}_2$ را نوشته (تمرین ۲۰.۶) و از آن‌جا $|Aut\ (\mathcal{R}_2 \oplus \mathcal{R}_2)|$ را پیدا کنید. نشان دهید تعداد کدهای متمایز معادل با $\mathcal{C}$ در $\mathcal{V}$ برابر $n!/|\mathrm{Aut}(\mathcal{C})|$ است و تمامی این کدها را برای $\mathcal{C} = \mathcal{R}_2 \oplus \mathcal{R}_2$ تعیین کنید.

تمرین ۱۴.۷ نشان دهید $\mathcal{H}_۷$ و $PG(۲,۲)$ هر دو دارای یک گروه اتومورفیسم ایزمورف با مجموعه ماتریس‌های ۳×۳ معکوس‌پذیر روی $F_۲$، یعنی $GL(۳,۲)$، هستند. مقدار $|\mathrm{Aut}(\mathcal{H}_۷)|$ و تعداد کدهای متمایز $\mathcal{C} \subseteq F_۲^۷$ معادل با $\mathcal{H}_۷$ را پیدا کنید. نتایج مشابه را برای $\mathcal{H}_n$، $n = ۲^c - ۱$، تعیین کنید.

تمرین ۱۵.۷ نشان دهید اگر $\mathcal{C}$ یک کد خطی دوتایی کامل t خطا تصحیح کننده به‌طول n باشد آنگاه محمل کدکلمات به وزن $d = ۲t + ۱$ در واقع قالب‌های یک سیستم اشتاینر $S(t+۱,d,n)$ روی $S = \{۱,۲,\ldots,n\}$ هستند. نتیجه بگیرید برای $۰ \le i \le t$، عدد $\binom{d-i}{t+۱-i}$ بخش‌کننده $\binom{n-i}{t+۱-i}$ است.

تمرین ۱۶.۷ تجزیه عدد $۴۰۹۶ = \binom{۹۰}{۲} + ۹۰ + ۱$ چه چیزی در رابطه با امکان وجود یک کد کامل دوتایی به‌طول ۹۰ می‌گوید؟ نشان دهید یک چنین کد کاملی وجود ندارد.

تمرین ۱۷.۷ نشان دهید اگر $\mathbf{u}$ و $\mathbf{v}$ بردارهای دوتایی باشند آنگاه

$$wt(\mathbf{u}+\mathbf{v}) = wt(\mathbf{u}) + wt(\mathbf{v}) - ۲c(\mathbf{u},\mathbf{v})$$

که $c(\mathbf{u},\mathbf{v})$ تعداد مختصات i است به‌قسمی‌که $u_i = v_i = ۱$. فرض کنید $G = (I_{۱۲}|P)$ که ۱۲ سطر و ستون P به‌وسیله رئوس یک بیست وجهی اندیس‌دار شده و $P_{ij} = ۰$ اگر و فقط اگر ij یک ضلع این بیست وجهی باشد. نشان دهید G یک ماتریس مولد برای یک $[۲۴,۱۲]$- کد خطی دوتایی خود—دوگان است و این‌که $G' (P|I_{۱۲})$ نیز مولد همین کد است. نشان دهید وزن هر کدکلمه مضربی از ۴ است ولی هیچ‌یک از آنها ۴ نیست و نتیجه بگیرید $\mathcal{C}$ دارای می‌نیمم فاصله ۸ است. نشان دهید کد پنچر شده $\mathcal{C}^\circ$ یک $[۲۳,۱۲]$- کد خطی کامل ۳ خطا تصحیح کننده است. (می‌توان نشان داد که $\mathcal{C}$ و $\mathcal{C}^\circ$ معادل با گلی کدهای $\mathcal{G}_{۲۴}$ و $\mathcal{G}_{۲۳}$ هستند.)

تمرین ۱۸.۷ فرض کنید H ماتریس کنترل تساوی ارائه شده در مثال ۳۵.۷ برای $[۴,۲]$-کد سه‌تایی همینگ بوده و فرض کنید

$$G = \begin{pmatrix} J+I & I & I \\ O & H & -H \end{pmatrix}$$

که I و J به‌ترتیب ماتریس‌های ۴×۴ همانی و تماماً ۱ می‌باشند. نشان دهید G یک ماتریس مولد برای یک $[۱۲,۶]$-کد سه‌تایی با می‌نیمم فاصله ۶ بوده، و این‌که کد پنچر شده $\mathcal{C}^\circ$ یک $[۱۱,۶]$- کد کامل ۲ خطا تصحیح کننده است. (می‌توان نشان داد که $\mathcal{C}$ و $\mathcal{C}^\circ$ به‌ترتیب معادل گلی کدهای $\mathcal{G}_{۱۲}$ و $\mathcal{G}_{۱۱}$ هستند.)

تمرین ۱۹.۷ کد رید — مولر[۱] $\mathcal{RM}(r,m)$، از مرتبه r و طول $n = ۲^m$، به‌صورت بازگشتی به‌شرح زیر تعریف می‌شود. کد $\mathcal{RM}(۰,m)$ برابر کد دوتایی تکراری به‌طول n بوده

[۱] The r-th order Reed - Muller code

و $\mathcal{RM}(m,m) = F_2^m$. اگر $0 < r < m$ آنگاه

$$\mathcal{RM}(r,m) = \mathcal{RM}(r,m-1) * \mathcal{RM}(r-1,m-1)$$

که * در تمرین ۲۰.۶ تعریف شده است. نشان دهید $\mathcal{RM}(r,m)$ یک کد خطی دوتایی به‌طول $n = 2^m$، بعد $k = \sum_{i=0}^{r} \binom{m}{i}$ و می‌نیمم فاصله $d = 2^{m-r}$ است. تمامی کدکلمه‌های کدهای $\mathcal{RM}(1,2)$ و $\mathcal{RM}(1,3)$ را بنویسید. یک ماتریس مولد G و از آن‌جا یک ماتریس کنترل تساوی H برای $\mathcal{RM}(1,3)$ بنویسید؛ با به‌کار بردن H ثابت کنید این کد دارای می‌نیمم فاصله ۴ است.

راهنمایی برای مطالعه بیشتر

مقاله کلاسیک ۱۹۴۸ شانون [Sh 48] به‌وسیله ویور (Weaver) با یک مقدمه فنی به‌صورت یک کتاب کوتاه [SW 63] چاپ شده است و کاملاً ارزش خواندن را دارد. اش (Ash) یک دید ریاضی دقیق و مفصل از نظریه اطلاعات ارائه می‌دهد [As 65]، درحالی‌که رضا [Re 61] یک دید مهندسی به موضوع دارد همچنان‌که مک‌الیس برخورد نسبتاً پیچیده‌ای را با نظریه کدگذاری و اطلاعات در [McE 77] ارائه می‌دهد. شامبرز [Ch 85] و جانز [Jo 79] یک مقدمه‌ای دقیق از نظریه اطلاعات را بیشتر از دید کاربردی مطرح می‌کنند درحالی‌که ولش [We 88] تاکید روی ارتباط با رمزنگاری دارد.

در نظریه کدگذاری، هیل [Hi 86] و پلس [Pl 82] هر دو روی توسعه موضوع کار را بیشتر از ما ادامه می‌دهند اگرچه از نظر شروع هم‌سطح هستیم. کتاب‌های درسی نسبتاً پیشرفته‌تری به‌وسیله برلیکمپ [Be 86]، بلیک ومولین [BM75، BM76]، پریتزل [Pr 92] و ون لنت [Li 82] نوشته شده‌اند درحالی‌که کتاب‌های مرجع در نظریه کدگذاری کارهای دائرةالمعارف‌گونه انجام شده به‌وسیله مک‌ویلیام و سلون [MS 77] و به‌وسیله پلس و هافمن [PH 98] می‌باشند. تامپسون [Th 83] یک دید تاریخی جالبی از شروع نظریه کدگذاری و بالاخص کدهای همینگ و گلی و ارتباط آنها با کره‌چینی و گروه‌های ساده ارائه می‌دهد. اندرسون [An 74] یک مقدمه خوبی راجع به پیشینه ترکیبیاتی لازم برای این ارتباطات، شامل سیستم اشتاینر $S(۵, ۸, ۲۴)$، نوشته درحالی‌که کانوی و سلون [CS 92] یک تحلیل خیلی عمیق‌تر و مفصل‌تر راجع به این موضوعات فراهم کرده‌اند. روابط بین کدها، گراف‌ها و طرح‌های قالبی به‌تفصیل توسط کامرون و ون لنت [CL 91] معرفی شده‌اند. کاربردهای هندسه جبری در کدگذاری به‌وسیله پریتزل [Pr 92]، گوپا [Go 88]، و ون لنت و ون درگیر [LG 88] مطالعه شده است.

کدهای با طول متغیر، همچنان‌که در فصل ۱ مطالعه شد، را می‌توان به‌عنوان اشیاء صرفاً جبری درنظر گرفت. مجموعه T^* متشکل از تمامی کلمات روی یک الفبای T یک تکواره[۱]

[۱] monoid

است، به این معنی که مجهز به یک عمل دوتایی (اتصال) است که تحت آن دارای عضو خنثی (کلمه خالی ε) بوده و خاصیت شرکت‌پذیری، $u(vw) = (uv)w$، دارد. یکتا دکدپذیری یک کد $\mathcal{C} \subseteq T^*$ معادل با این شرط است که $\mathcal{C}$ باید مجموعه‌ای از مولدهای آزاد برای زیرتکواره تولید شده به‌وسیله $\mathcal{C}$ باشد. اینها و دیگر روابط مهم بین کدها و جبر با تفصیل زیاد به‌وسیله برستل و پیرن [BP 85] بیان شده‌اند.

درخت‌ها را در فصل ۱ برای توصیف دسته‌های خاصی از کدها، همچون کدهای لحظه‌ای، معرفی کردیم. درخت‌ها هم در نظریه گراف مهم هستند و هم در زمینه‌های دیگری همچون علوم کامپیوتر. الگوریتم هافمن تنها یکی از الگوریتم‌های درختی بحث شده در [Kn 73] است. برای دیگر کاربردهای الگوریتم هافمن به [De 74]، [Ev 79]، [Kn 73]، [ST 81]، [Zi 59] مراجعه کنید.

آنتروپی، معرفی شده در فصل ۳، یک نقش مهمی نیز در ترمودینامیک به‌عنوان اندازه‌ای برای بی‌نظمی یک سیستم دارد که در آن p_i احتمال این است که سیستم در i امین حالت ممکن خود باشد. قانون دوم ترمودینامیک می‌گوید آنتروپی کاهش پیدا نمی‌کند، از این‌رو با نشان دادن این‌که سیستم‌ها تمایل در جهت بی‌نظمی دارند یک جهت برای زمان فراهم می‌شود. برلوین[۱] روابط بین نظریه اطلاعات و ترمودینامیک را در [Br 56] بحث می‌کند. همچنین روابطی قوی بین آنتروپی و نظریه ارگدیک[۲] (نظریه تبدیلات حافظ اندازه) وجود دارد؛ برای نمونه به [Bi 65] مراجعه کنید.

نظریه احتمال لازم برای مطالعه این کتاب در اکثر کتاب‌های احتمال پوشانده شده است. قانون اعداد بزرگ که در اثبات قضیه اساسی شانون به‌کار می‌رود در ضمیمه B ارائه شده است؛ برای اثبات و یک بحث کامل‌تر مرجع [Fe 50] را توصیه می‌کنیم. مشابهاً، کتاب‌های زیادی در جبر خطی موجود هستند که حاوی مطالب مورد نیاز ما در فصول ٦ و ٧ می‌باشند؛ [BR 98] یک مثال خوب است. ما از چند نتیجه در آنالیز و حساب، مانند قضیه مقدار میانی و تقریب استرلینگ برای $n!$، نیز استفاده می‌کنیم؛ فیشر [Fi 83] و لنگ [La 83] مراجع خوبی در این رابطه در سطح کارشناسی می‌باشند. میدان‌های متناهی در فصل‌های ٦ و ٧ به‌کار می‌روند؛ برای زمینه و کاربردهای بیشتر می‌توان به [KR 83] مراجعه کرد.

در ارتباط با خواندن مقالات اصلی به‌منظور فهم چگونه اندیشیدن بنیان‌گذاران یک موضوع حرف‌های زیادی می‌توان زد. این بالاخص در زمینه نظریه اطلاعات و کدگذاری صادق است: بسیاری از این مقالات به‌آسانی در دسترس بوده و بدون یک زمینه ریاضی قوی قابل مطالعه هستند. مجموعه مقالات مهم جمع‌آوری شده به‌وسیله برلیکمپ [Be 74] و سلپین [Sl 74] حاوی اکثر پیشرفت‌های مهم این موضوع در ۲۵ سال اول می‌باشند. خوانندگانی که دسترسی به مجلاتی چون Bell System Technical Journal،

---

[۱] Brillouin [۲] Ergodic theory

IEEE Transactions in Information Theory، و Information and Control دارند مقالات تحقیقی جالب و مهمی را خواهند یافت؛ به‌عنوان نمونه به مقاله شانون [Sh 48]، یک مقاله مهم راجع به کدهای با طول متغیر به‌وسیله گیلبرت و مور [GM 59]، نتایجی راجع به غیر یکتایی کدهای هافمن به‌وسیله شوارتز [Sc 64] و گلمب [Go 80]، و مقاله‌ای راجع به قماربازی بوسیله کیلی [Ke 56]، که تمرینات ۱۱.۵ و ۱۲.۵ از روی آن ساخته شده‌اند، می‌توان اشاره کرد.

ضمیمه A
اثبات قضیه سردناس — پترسن

قضیه سردناس — پترسن[1] بدون اثبات کامل به‌عنوان قضیه ۱۰.۱ در بخش ۲.۱ بیان شد. به‌خاطر بیاورید که یک کد $\mathcal{C} \subseteq T^*$ یکتا دکدپذیر تعریف می‌شود اگر هر وقت $u_1 \cdots u_l = v_1 \cdots v_m$ که $u_i, v_j \in \mathcal{C}$، آنگاه $l = m$ و برای هر i تساوی $u_i = v_i$ برقرار باشد. برای یک کد مفروض $\mathcal{C}$ تعریف می‌کنیم $\mathcal{C}_0 = \mathcal{C}$ و برای هر $n \geq 1$

$$\mathcal{C}_n = \{w \in T^+ \mid v \in \mathcal{C}, u \in \mathcal{C}_{n-1} \text{ یا } v \in \mathcal{C}_{n-1}, u \in \mathcal{C} \text{ که } uw = v\}$$

و $\mathcal{C}_\infty = \bigcup_{n=1}^{\infty} \mathcal{C}_n$. در این‌صورت قضیه می‌گوید $\mathcal{C}$ یکتا دکدپذیر است اگر و فقط اگر $\mathcal{C} \cap \mathcal{C}_\infty = \emptyset$. اثباتی را که در این‌جا ارائه می‌دهیم براساس مراجع [Ba 63] و [Se 67] است.

ابتدا به چند نماد نیاز داریم. اگر $u, v \in T^*$ و u یک پیشوند v باشد، یعنی به‌ازای عضوی چون $w \in T^*$ تساوی $v = uw$ برقرار باشد، آنگاه می‌نویسیم $u \leq v$؛ همچنین از $w = u^{-1}v$ برای بیان این‌که w از حذف پیشوند u در v حاصل می‌شود استفاده می‌کنیم. (توجه کنید که u^{-1} به تنهایی تعریف نمی‌شود.) افزون‌براین، اگر $u \neq v$ آنگاه می‌نویسیم $u < v$. حال می‌توانیم اثبات را شروع کنیم.

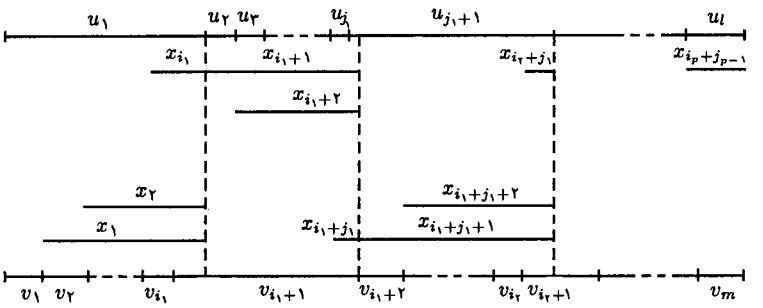
($\Longleftarrow$) اگر $\mathcal{C}$ یکتا دکدپذیر نباشد آنگاه می‌توان یک کددنباله با طول می‌نیمم، مثلاً

$$u_1 \cdots u_l = v_1 \cdots v_m, \qquad (*)$$

انتخاب کرد که $u_i, v_j \in \mathcal{C}$ و اگر $l = m$ آنگاه برای مقداری چون i نامساوی $u_i \neq v_i$ برقرار نباشد. به‌جهت می‌نیمال بودن دنباله داده شده به‌وسیله (*) داریم $u_1 \neq v_1$ (در غیر این‌صورت $u_2 \cdots u_l = v_2 \cdots v_m$ یک کددنباله مبهم کوتاه‌تری خواهد بود)، از این‌رو (*)

[1] Sardinas - Patterson theorem

ایجاب می‌کند $u_1 < v_1$ یا $v_1 < u_1$؛ می‌توان فرض کرد $v_1 < u_1$. در این‌صورت کلمه غیرخالی $x_1 = v_1^{-1}u_1$ در C_1 قرار دارد زیرا $u_1 = v_1x_1$ و $u_1, v_1 \in C_0$. اگر $v_1v_2 < u_1$، آنگاه چون $v_2 \in C$ و $x_1 = v_1^{-1}u_1 = v_2(v_1v_2)^{-1}u_1$، کلمه $x_2 = (v_1v_2)^{-1}u_1$ در C_2 است. این روش را تا رسیدن به بزرگ‌ترین عدد i_1 با خاصیت $v_1 \dots v_{i_1} < u_1$ ادامه می‌دهیم؛ توجه کنید که برای $1 \le n \le i_1$ کلمه $x_n = (v_1 \dots v_n)^{-1}u_1$ در C_n قرار دارد. شکل ۱.A این وضعیت را توضیح می‌دهد.



شکل ۱.A

در مرحله بعد باید داشته باشیم $u_1 \le v_1 \dots v_{i_1+1}$. اگر $u_1 = v_1 \dots v_{i_1+1}$ (به‌قسمی‌که از می‌نیمال بودن نتیجه شود $l = 1$ و $m = i_1 + 1$)، آنگاه $v_{i_1+1} = x_{i_1} \in C \cap C_{i_1} \subseteq C \cap C_\infty$، و از این‌رو $C \cap C_\infty \neq \emptyset$. پس می‌توان فرض کرد $u_1 < v_1 \dots v_{i_1+1}$. در این‌صورت $x_{i_1+1} = u_1^{-1}v_1 \dots v_{i_1+1}$ در C_{i_1+1} قرار دارد چون $x_{i_1}x_{i_1+1} = v_{i_1+1} \in C$ و $x_{i_1} \in C_{i_1}$. اگر $u_1u_2 < v_1 \dots v_{i_1+1}$، آنگاه کلمه $x_{i_1+2} = (u_1u_2)^{-1}(v_1 \dots v_{i_1+1})$ در C_{i_1+2} قرار دارد زیرا $u_2x_{i_1+2} = x_{i_1+1} \in C_{i_1+1}$ و $u_2 \in C$. مجدداً با ادامه این روش به بزرگ‌ترین عدد صحیح j_1 می‌رسیم به‌قسمی‌که $u_1 \dots u_{j_1} < v_1 \dots v_{i_1+1}$، و از این‌رو

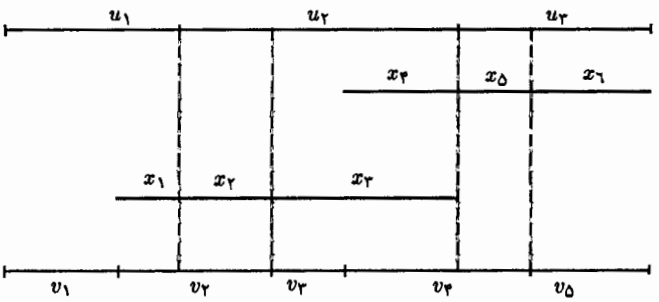
$$x_{i_1+j_1} = (u_1 \dots u_{j_1})^{-1}(v_1 \dots v_{i_1+1}) \in C_{i_1+j_1}.$$

حال $v_1 \dots v_{i_1+1} \le u_1 \dots u_{j_1+1}$، و اگر تساوی برقرار باشد آنگاه (همچون قبل) به انتهای کددنباله می‌نیمال مبهم (*) رسیده‌ایم و $u_{j_1+1} = x_{i_1+j_1} \in C \cap C_\infty$. با فرض $u_1 \dots u_{j_1+1} < v_1 \dots v_{i_1+1}$، کلمه $x_{i_1+j_1+1} = (v_1 \dots v_{i_1+1})^{-1}(u_1 \dots u_{j_1+1})$ در $C_{i_1+j_1+1}$ قرار دارد زیرا $x_{i_1+j_1}x_{i_1+j_1+1} = u_{j_1+1} \in C$ با $x_{i_1+j_1} \in C_{i_1+j_1}$. با ادامه این روش به بزرگ‌ترین عدد صحیح i_2 می‌رسیم به‌قسمی‌که $v_1 \dots v_{i_2} < u_1 \dots u_{j_1+1}$. نتیجه می‌شود $x_{i_2+j_1} = (v_1 \dots v_{i_2})^{-1}(u_1 \dots u_{j_1+1}) \in C_{i_2+j_1}$.

در مرحله بعد $u_1 \dots u_{j_1+1} \le v_1 \dots v_{i_2+1}$، و همچون قبل کار را تا رسیدن به $u_1 \le v_1 \dots v_{i_1+1}$ ادامه می‌دهیم. نهایتاً باید به انتهای کددنباله (*) برسیم. حال $|u_l| \neq |v_m|$، زیرا در غیر این‌صورت $u_l = v_m$ و از این‌رو $u_1 \dots u_{l-1} = v_1 \dots v_{m-1}$

که در تناقض با می‌نیمال بودن است. اگر $|u_l| > |v_m|$ (همچون شکل ۱.A) آنگاه به عددی چون p خواهیم رسید به‌قسمی‌که $m-1=i_p$ و $v_m = x_{i_p+j_{p-1}} \in \mathcal{C}\cap\mathcal{C}_\infty$. از طرف دیگر، اگر $|u_l| < |v_m|$ آنگاه به عددی چون p می‌رسیم که $l-1=j_p$ و $u_l = x_{i_p+j_p} \in \mathcal{C}\cap\mathcal{C}_\infty$.

مثال ۱.A فرض کنید دنباله مبهم می‌نیمال (*) به‌فرم $u_1u_2u_3 = v_1v_2v_3v_4v_5$ باشد، که روی‌هم بودن کلمات u_i و v_j مطابق شکل ۲.A می‌باشد.



شکل ۲.A

با دنبال کردن روش بالا ملاحظه خواهیم کرد $i_1=1$، $j_1=1$، $i_2=3$، $j_2=2$، $i_3=4=m-1$، پس $p=3$ و $v_5 = x_6 \in \mathcal{C}\cap\mathcal{C}_6 \subseteq \mathcal{C}\cap\mathcal{C}_\infty$.

($\Longrightarrow$) فرض کنید $\mathcal{C}\cap\mathcal{C}_\infty \neq \emptyset$، پس برای عددی چون $n \geq 1$ خواهیم داشت $\mathcal{C}\cap\mathcal{C}_n \neq \emptyset$؛ فرض کنید v_n عضوی از $\mathcal{C}\cap\mathcal{C}_n$ باشد. با به‌کار بردن تعریف مجموعه‌های $\mathcal{C}_n$،...،$\mathcal{C}_2$ ملاحظه می‌کنیم گزاره زیر برای $2 \leq k \leq n$ درست است:

(S_k) برای کلماتی چون $u_{k-1} \in \mathcal{C}$ و $v_{k-1} \in \mathcal{C}_{k-1}$ خواهیم داشت $u_{k-1}v_k = v_{k-1}$ یا $v_{k-1}v_k = u_{k-1}$.

مشابهاً از تعریف $\mathcal{C}_1$ نتیجه می‌شود

(S_1) به‌ازای کلماتی چون $u, u' \in \mathcal{C}$ خواهیم داشت $uv_1 = u'$.

در این‌جا همواره $v_k \neq \varepsilon$، و بالاخص $u \neq u'$، حقیقتی که بعداً به آن نیاز داریم. گزاره‌های (S_k) ما را قادر به ساختن کلمه‌ای می‌کنند که به دو طریق قابل تجزیه به کدکلمه‌ها خواهد بود. برای این منظور نیاز به این داریم که نشان دهیم برای هر $k \in \{1,\ldots,n-1\}$ عضو $v_k \in \mathcal{C}_k$ نیز در گزاره زیر صدق می‌کند:

(T_k) برای عناصری چون $y_k, z_k \in \mathcal{C}^*$ رابطه $v_ky_k = z_k$ برقرار است.

اولاً، (T_{n-1}) درست است زیرا از (S_n) نتیجه می‌شود

$$v_{n-1}y_{n-1} = z_{n-1}$$

که در آن یا $y_{n-1} = \varepsilon$ و $z_{n-1} = u_{n-1}v_n$، یا $y_{n-1} = v_n$ و $z_{n-1} = u_{n-1}$؛ در هر حالت $y_{n-1}, z_{n-1} \in \mathcal{C}^*$ زیرا $u_{n-1}, v_n \in \mathcal{C}$. حال نشان خواهیم داد برای $2 \leq k \leq n-1$ از

(T_k) نتیجه می‌شود (T_{k-1}). فرض کنید (T_k) درست باشد. از گزاره (S_k) نتیجه می‌شود

$$u_{k-1}v_k = v_{k-1} \quad \text{یا} \quad v_{k-1}v_k = u_{k-1},$$

پس یا $v_{k-1}y_k = u_{k-1}v_ky_k = u_{k-1}z_k$ یا $v_{k-1}z_k = v_{k-1}v_ky_k = u_{k-1}y_k$.
این دو عبارت هر دو به‌شکل $v_{k-1}y_{k-1} = z_{k-1}$ هستند که به‌ترتیب خواهیم داشت $z_{k-1} = u_{k-1}z_k$، $y_{k-1} = y_k$ یا $y_{k-1} = z_k$، $z_{k-1} = u_{k-1}y_k$. در هر حالت y_{k-1} و z_{k-1} اعضای C^* هستند (زیرا $y_k, z_k \in C^*$ و $u_{k-1} \in C$)، پس (T_{k-1}) ثابت شده است.
از این‌رو (T_k) همواره درست است، پس با انتخاب $k = 1$ خواهیم داشت $v_1y_1 = z_1$ که y_1 و z_1 اعضایی از C^* هستند. سپس از (S_1) نتیجه می‌شود $u'y_1 = uv_1y_1 = uz_1$، که $y_1, z_1 \in C^*$ و u و u' کدکلمه‌های متمایزی هستند. پس معادله $u'y_1 = uz_1$ منجر به دو تجزیه متمایز یک کددنباله شده، و از این‌رو C یکتا دکدپذیر نیست. ■

مثال A.۲ برای توضیح این قضیه به مثال ۱۲.۱ از بخش ۲.۱ باز می‌گردیم که $C = \{0, 1, 2, 10\}$. در آن‌جا ملاحظه کردیم $1 \in C \cap C_3$، پس براساس نمادگذاری بالا قرار می‌دهیم $n = 3$ و $v_3 = 1$. در این‌صورت گزاره (S_k) به‌فرم زیر می‌باشد

(S_3) $01 = 0.1$، یعنی این‌که $v_2v_3 = u_2$ که $u_2 = 01 \in C$ و $v_2 = 0 \in C_2$؛

(S_2) $10 = 1.0$، یعنی این‌که $u_1v_2 = v_1$ که $u_1 = 1 \in C$ و $v_1 = 10 \in C_1$؛

(S_1) $210 = 2.10$، یعنی این‌که $uv_1 = u'$ که $u = 2 \in C$ و $u' = 210 \in C$.

پس نتیجه (T_2)، یعنی این‌که $v_2y_2 = z_2$، که تبدیل به $01 = 0.1$ می‌شود که $y_2 = 1 \in C^*$ و $z_2 = 01 \in C^*$. مشابهاً (T_1)، یعنی این‌که $v_1y_1 = z_1$، که تبدیل به $101 = 10.1$ با $y_1 = 1 \in C^*$ و $z_1 = 101 = 1.01 \in C^*$ می‌شود. با به‌کار بردن (S_1) و (T_1) و تجزیه z_1 داریم

$$210.1 = 2.10.1 = 2.101 = 2.1.01.$$

این منجر به دو تجزیه 210.1 و $2.1.01$ از دنباله 2101 می‌شود که یکتا دکدناپذیری C را تایید می‌کند.

ضمیمه B
قانون اعداد بزرگ

در اثبات قضیه اساسی شانون (قضیه ۹.۵) نیاز به یک برآورد از تعداد خطای ایجاد شده در ارسال کدکلمه به‌طول n $\mathbf{u} = u_1 \dots u_n$ می‌باشد. در حالت یک کانال BSC، که $A = B = \mathbf{Z}_2$، بسته به این‌که u_i درست یا نادرست ارسال شده باشد داریم $e_i = 0$ یا $e_i = 1$. این دو پیشامد مستقل از این‌که روی دیگر ارقام $\mathbf{u}$ چه اتفاق افتاده است دارای احتمال‌های P و $Q(= \overline{P})$ هستند، پس می‌توان $e_1, \dots, e_n$ را به‌عنوان خروجی n آزمایش متوالی برنولی (متغیرهای تصادفی مستقل هم‌توزیع) درنظر گرفت. اگر در هر یک از e_i ها مقادیر ۰ و ۱ را به‌عنوان یک عدد حقیقی درنظر بگیریم آنگاه تعداد خطا برابر $\sum_i e_i$ خواهد بود. قانون اعداد بزرگ راجع به مجموع (معادلاً میانگین) مقادیر تعداد زیادی از آزمایش‌های برنولی صحبت می‌کند، از این‌رو برآوردی از تعداد خطا ارائه می‌دهد.

فرض کنید X یک متغیر تصادفی بوده که تعداد متناهی مقدار حقیقی x_j را با احتمال p_j اختیار می‌کند، پس $0 \leq p_j \leq 1$ و $\sum_j p_j = 1$. مقدار میانگین X برابر است با

$$\mu = E(X) = \sum_j p_j x_j.$$

حال فرض کنید $X_1, \dots, X_n$ آزمایش‌های متوالی برنولی از X باشند، یعنی این‌که n متغیر تصادفی مستقل که مشابه X عمل می‌کنند. (به‌عنوان مثال می‌توان پرتاب‌های مکرر یک تاس یا یک سکه را درنظر گرفت.) اگر

$$Y = \frac{1}{n} \sum_{i=1}^{n} X_i$$

میانگین n خروجی باشد، آنگاه شهوداً انتظار داریم که برای n های بزرگ مقدار Y به μ

نزدیک باشد. به‌عنوان مثال، اگر X یک سکه سالم باشد، و ما مقادیر $x_j = ۱$ و $x_j = ۰$ را به طرفین سکه نسبت دهیم آنگاه $\mu = \frac{۱}{۲}$ و همچنین توقع داریم که $Y \approx \frac{۱}{۲}$.

البته نمی‌توانیم تضمین کنیم که در تمامی حالات داشته باشیم $Y \approx \mu$. اگر سکه را $n = ۱۰$ بار پرتاب کنیم آنگاه ظاهر شدن ده مقدار $x = ۱$ ($Y = ۱$) بعید به‌نظر می‌رسد، ولی غیرممکن نیست: احتمال آن $۰.۰۰۱ \approx ۱/۱۰۲۴ = ۲^{-۱۰}$ است که خیلی کوچک بوده ولی صفر نیست. حتی خروجی ۶ بار $x = ۱$ از ۱۰ پرتاب ($Y = ۰.۶$) خیلی عجیب نیست زیرا احتمال این پیشامد $۰.۲۰۵ \approx ۲۱۰/\binom{۱۰}{۶}$ است در حالی‌که محتمل‌ترین پیشامد، ظاهر شدن ۵ بار $x = ۱$ از بین ۱۰ پرتاب، احتمالی حدود ۰.۲۴۶ دارد. اگر سکه را $n = ۱۰۰$ بار پرتاب کنیم آنگاه بسیار بعید است که Y مقداری نزدیک به $\frac{۱}{۲}$ داشته باشد: به‌عنوان مثال در این حالت احتمال $Y = ۱$ برابر $۱۰^{-۳۰} \approx ۲^{-۱۰۰}$، و احتمال $Y = ۰.۶$ برابر $۰.۰۱ \approx ۲۱۰۰/\binom{۱۰۰}{۶۰}$ می‌باشد، پس هر دو پیشامد بسیار غیرمتحمل هستند اگرچه غیرممکن نیستند.

قانون اعداد بزرگ این درک شهودی را تایید می‌کند. براساس این قانون با افزایش n، احتمال نزدیک بودن Y به μ افزایش می‌یابد. به بیانی دقیق‌تر این قانون می‌گوید برای هر $\eta > ۰$ با افزایش $n \longrightarrow \infty$ احتمال نامساوی $|Y - \mu| \leq \eta$ به ۱ میل می‌کند، معادلاً

$$\lim_{n \longrightarrow \infty} Pr\left(\left|\frac{۱}{n}\sum_{i=۱}^{n} X_i - \mu\right| > \eta\right) = ۰.$$

این در واقع بیشتر به نام قانون ضعیف اعداد بزرگ معروف است زیرا نسخه‌های قوی‌تری از این قانون وجود دارند. برای جزییات بیشتر در این رابطه و قضایای حدی دیگر در آمار به [Fe50] مراجعه کنید.

ضمیمه C
اثبات قضیه اساسی شانون

در بخش ۴.۵ قضیه اساسی شانون را برای کانال BSC بیان کردیم.

قضیه ۹.۵ فرض کنید Γ یک کانال دوتایی متقارن با $P > \frac{1}{2}$ باشد، پس ظرفیت کانال برابر $C = 1 - H(P) > 0$ است. نیز فرض کنید ε و δ اعداد مثبتی باشند. در این‌صورت برای تمامی اعداد به‌قدر کافی بزرگ n، یک کد $\mathcal{C} \subseteq F_2^n$ با نرخ R وجود دارد به‌قسمی‌که $C - \varepsilon \leq R < C$ و تحت کدگشایی نزدیک‌ترین همسایه رابطه $Pr_E < \delta$ برقرار است.

اکنون با پر کردن شکاف‌های موجود در اثبات معرفی شده در بخش ۴.۵ یک اثبات کامل ارائه می‌شود.

اثبات فرض کنید $\mathcal{V} = \mathbf{Z}_2^n$. ما یک کد $\mathcal{C} \subseteq \mathcal{V}$ را به‌عنوان یک دنباله مرتب $(\mathbf{u}_1, \ldots, \mathbf{u}_M)$ از اعضای متمایز $\mathcal{V}$ درنظر می‌گیریم، پس آرایش‌های متفاوت از اعضای یکسان به‌عنوان کدهای مختلفی درنظر گرفته می‌شوند. این تنها یک تدبیر فنی برای اثبات قضیه است، پس از اثبات این‌که یک کد مرتب در شرایط قضیه صدق می‌کند می‌توانیم ترتیب را فراموش کنیم.

ابتدا کدگشایی را درنظر می‌گیریم. اجازه دهید یک عدد کوچک $\eta > 0$ انتخاب کرده (بعداً میزان کوچکی آن را مشخص خواهیم کرد)، و قرار دهیم $\rho = n(Q + \eta)$ که $Q = \overline{P}$. انگیزه این کار این است که در هر کدکلمه به‌طول n انتظار nQ خطا را داریم، پس کلمات ارسال‌شده و دریافت‌شده $\mathbf{u}$ و $\mathbf{v}$ احتمالاً در شرط $d(\mathbf{u}, \mathbf{v}) \approx nQ$ صدق می‌کنند، با درنظر گرفتن ρ کمی بزرگ‌تر از nQ می‌توان با احتمال بالا انتظار $d(\mathbf{u}, \mathbf{v}) \leq \rho$ را داشت.

از ρ برای پیدا کردن یک کران بالا برای میانگین احتمال خطا Pr_E استفاده خواهیم کرد. فرض کنید یک کدکلمه $\mathbf{u}_i \in \mathcal{C}$ ارسال شده، و $\mathbf{v} = \mathbf{u}_i + \mathbf{e} \in \mathcal{V}$ دریافت شده باشد که $\mathbf{e}$ بردار تصادفی خطا می‌باشد. اگر $d(\mathbf{u}_i, \mathbf{v}) \leq \rho$ و برای هر $j \neq i$ داشته باشیم $d(\mathbf{u}_j, \mathbf{v}) > \rho$، آنگاه از کدگشایی نزدیک‌ترین همسایه نتیجه می‌شود $\Delta(\mathbf{v}) = \mathbf{u}_i$ که در واقع درست است.

معادلاً، اگر کدگشایی درست نباشد آنگاه یا $d(\mathbf{u}_i, \mathbf{v}) > \rho$ و یا برای عددی چون $j \neq i$ رابطه $d(\mathbf{u}_j, \mathbf{v}) \leq \rho$ برقرار است. با میانگین گرفتن روی تمامی بردارهای خطای e، نتیجه می‌گیریم احتمال شرطی $Pr(\Delta(\mathbf{v}) \neq \mathbf{u}_i \mid \mathbf{u}_i)$، احتمال کدگشایی نادرست به‌شرط ارسال $\mathbf{u}_i$، در شرط زیر صدق می‌کند

$$Pr(\Delta(\mathbf{v}) \neq \mathbf{u}_i \mid \mathbf{u}_i) \leq Pr(d(\mathbf{u}_i, \mathbf{u}_i + \mathbf{e}) > \rho) + \sum_{j \neq i} Pr(d(\mathbf{u}_j, \mathbf{u}_i + \mathbf{e}) \leq \rho). \quad (C.1)$$

نشان می‌دهیم که جمله اول سمت راست این رابطه را می‌توان به‌قدر دلخواه کوچک کرد. با نوشتن $\mathbf{e} = (e_1, \ldots, e_n)$ که $e_i \in \{0, 1\}$، داریم

$$d(\mathbf{u}_i, \mathbf{u}_i + \mathbf{e}) = wt(\mathbf{e}) = \sum_{k=1}^{n} e_k$$

که عمل جمع در $\mathbf{Z}$ صورت می‌گیرد و نه در $\mathbf{Z}_2$. حال از $\rho = n(Q + \eta)$ نتیجه می‌شود

$$\begin{aligned} Pr\left(d(\mathbf{u}_i, \mathbf{u}_i + \mathbf{e}) > \rho\right) &= Pr\left(\frac{1}{n}\sum_{k=1}^{n} e_k > Q + \eta\right) \\ &\leq Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n} e_k - Q\right| > \eta\right). \end{aligned}$$

می‌توان $e_1, \ldots, e_n$ را به‌عنوان آزمایش‌های برنولی درنظر گرفت که مقادیر 0 و 1 را به‌ترتیب با احتمال‌های P و Q اختیار می‌کنند. بنابراین میانگین $\mu = E(e_k)$ برای هر e_k برابر $Q.1 + P.0 = Q$ است، پس از قانون ضعیف اعداد بزرگ (ضمیمه B) نتیجه می‌شود

$$Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n} e_k - Q\right| > \eta\right) \longrightarrow 0 \quad \text{آنگاه} \quad n \longrightarrow \infty \quad \text{اگر}.$$

(این بدین معنی است که برای مقادیر بزرگ n، میانگین دنباله $e_1, \ldots, e_n$ احتمالاً نزدیک به $Q = \mu(e)$ است.) پس

$$Pr\left(\left|\frac{1}{n}d(u_i, u_i + e) - Q\right) | > \rho\right) \longrightarrow 0 \quad \text{آنگاه} \quad n \longrightarrow \infty \quad \text{اگر}\ ،$$

از این‌رو برای تمامی مقادیر به‌قدر کافی بزرگ n داریم

$$Pr\left(d(\mathbf{u}_i, \mathbf{u}_i + \mathbf{e}) > \rho\right) < \frac{\delta}{2}. \quad (C.2)$$

این مربوط به اولین جمله در (C.۱) می‌شود. با میانگین گرفتن از (C.۱) روی تمامی کدکلمات $\mathbf{u}_i \in \mathcal{C}$ (با فرض هم‌احتمال بودن آنها)، می‌بینیم که احتمال خطای $\mathcal{C}$ در رابطه زیر صدق می‌کند

$$\begin{aligned} Pr_E &= \frac{1}{M}\sum_{i=1}^{M} Pr(\Delta(\mathbf{v}) \neq \mathbf{u}_i \mid \mathbf{u}_i) \\ &\leq \frac{1}{M}\sum_{i=1}^{M}\left(Pr(d(\mathbf{u}_i, \mathbf{u}_i+\mathbf{e}) > \rho) + \sum_{j\neq i} Pr(d(\mathbf{u}_j, \mathbf{u}_i+\mathbf{e}) \leq \rho)\right) \\ &< \frac{1}{M}\sum_{i=1}^{M}\left(\frac{\delta}{2} + \sum_{j\neq i} Pr(d(\mathbf{u}_j, \mathbf{u}_i+\mathbf{e}) \leq \rho)\right) \\ &= \frac{\delta}{2} + \frac{1}{M}\sum_{i=1}^{M}\sum_{j\neq i} Pr(d(\mathbf{u}_j, \mathbf{u}_i+\mathbf{e}) \leq \rho). \end{aligned} \tag{C.۳}$$

کار کردن با جمع مضاعف در (C.۳) دشوار است زیرا وابستگی شدید به کد انتخاب شده $\mathcal{C} = (\mathbf{u}_1, \ldots, \mathbf{u}_M)$ دارد: بسته به دور و نزدیک بودن کلمات از یکدیگر این احتمال‌ها بزرگ یا کوچک می‌باشند. ایده برجسته شانون خارج شدن از این وابستگی به‌وسیله میانگین گرفتن روی تمامی انتخاب‌های ممکن برای $\mathcal{C}$ می‌باشد. اگر $f(\mathcal{C})$ یک عدد نسبت داد شده به کد $\mathcal{C}$ باشد، آنگاه میانگین f برابر است با

$$\bar{f} = \frac{(2^n - M)!}{2^n!}\sum_{\mathcal{C}} f(\mathcal{C}),$$

که عمل جمع روی تمامی $2^n!/(2^n - M)!$ کدهای مرتب M عضوی متمایز $\mathcal{C} = (\mathbf{u}_1, \ldots, \mathbf{u}_M)$ در $\mathcal{V}$ انجام می‌شود. (نباید ابهامی در رابطه با نماد $\bar{f} = 1 - f$، که در این‌جا از آن استفاده نمی‌کنیم، وجود داشته باشد.) با اعمال این روی (C.۳)، ملاحظه می‌کنیم که میانگین احتمال خطا برای این کدها در رابطه زیر صدق می‌کند

$$\begin{aligned} \overline{Pr_E} &< \frac{\delta}{2} + \frac{1}{M}\sum_i \sum_{j\neq i} \overline{Pr(d(\mathbf{u}_j, \mathbf{u}_i+\mathbf{e}) \leq \rho)} \\ &= \frac{\delta}{2} + \frac{1}{M}\sum_i \sum_{j\neq i} \overline{Pr(d(\mathbf{u}_j, \mathbf{u}_i+\mathbf{e}) \leq \rho)}; \end{aligned} \tag{C.۴}$$

در خط اول از این واقعیت که δ و M از $\mathcal{C}$ مستقل هستند، و در خط دوم از این‌که میانگین یک جمع برابر جمع میانگین آنها است استفاده می‌کنیم.

برای کار کردن با (C.۴)، اجازه دهید یک زوج پانویس $i \neq j$ انتخاب کنیم. حال $d(\mathbf{u}_j, \mathbf{u}_i + \mathbf{e}) = d(\mathbf{u}_j - \mathbf{u}_i, \mathbf{e})$، پس $d(\mathbf{u}_j, \mathbf{u}_i + \mathbf{e}) \leq \rho$ اگر و فقط اگر $\mathbf{u}_j - \mathbf{u}_i$ در $S_\rho(\mathbf{e})$، کره به مرکز $\mathbf{e}$ و شعاع ρ، قرار داشته باشد. فرض کنید

$$f_{\mathbf{e}}(\mathcal{C}) = \begin{cases} ۱, & \mathbf{u}_j - \mathbf{u}_i \in S_\rho(\mathbf{e}), \\ ۰, & \mathbf{u}_j - \mathbf{u}_i \notin S_\rho(\mathbf{e}). \end{cases}$$

در این صورت

$$\overline{Pr(d(\mathbf{u}_j, \mathbf{u}_i + \mathbf{e}) \leq \rho)} = \overline{Pr(\mathbf{u}_j - \mathbf{u}_i \in S_\rho(\mathbf{e}))}$$

$$= \frac{(۲^n - M)!}{۲^n!} \sum_{\mathcal{C}} Pr(\mathbf{u}_j - \mathbf{u}_i \in S_\rho(\mathbf{e}))$$

$$= \frac{(۲^n - M)!}{۲^n!} \sum_{\mathcal{C}} \left(\sum_{\mathbf{e}} Pr(\mathbf{e}) f_{\mathbf{e}}(\mathcal{C}) \right)$$

$$= \sum_{\mathbf{e}} \left(Pr(\mathbf{e}) \frac{(۲^n - M)!}{۲^n!} \sum_{\mathcal{C}} f_{\mathbf{e}}(\mathcal{C}) \right)$$

$$= \sum_{\mathbf{e}} Pr(\mathbf{e}) \overline{f_{\mathbf{e}}(\mathcal{C})}, \qquad (C.۵)$$

که $Pr(\mathbf{e})$ برابر احتمال یک بردار خطای $\mathbf{e}$ به وزن w است یعنی $Pr(\mathbf{e}) = P^{n-w}Q^w$. (این بحث صرفاً نشان می‌دهد که اعمال میانگین‌گیری روی $\mathcal{C}$ و $\mathbf{e}$ با هم جابه‌جا می‌شوند.) حال برای هر بردار خطای ثابت $\mathbf{e}$، $\overline{f_{\mathbf{e}}(\mathcal{C})}$ برابر درصد کدهایی چون $\mathcal{C}$ است که $\mathbf{u}_j - \mathbf{u}_i \in S_\rho(\mathbf{e})$. با تغییر $\mathcal{C} = (\mathbf{u}_۱, \ldots, \mathbf{u}_M)$ روی تمامی کدها، $\mathbf{u}_i$ و $\mathbf{u}_j$ روی تمامی زوج‌های متمایز در $\mathcal{V}$ تغییر می‌کنند، و هر یک چنین زوجی از احتمال ظاهر شدن یکسانی برخوردار است، پس $\mathbf{u}_j - \mathbf{u}_i$ روی $\mathcal{V}\backslash\{\mathbf{0}\}$ تغییر کرده و همه اعضای ناصفر $\mathcal{V}$ با احتمال یکسان ظاهر می‌شوند. نتیجه می‌شود که برای هر $\mathbf{e} \in \mathcal{V}$ داریم

$$\overline{f_{\mathbf{e}}(\mathcal{C})} = \frac{|S_\rho(\mathbf{e})\backslash\{\mathbf{0}\}|}{|\mathcal{V}\backslash\{\mathbf{0}\}|} \leq \frac{|S_\rho(\mathbf{e})|}{|\mathcal{V}\backslash\{\mathbf{0}\}|} = \frac{۱}{۲^n - ۱} \sum_{r \leq \rho} \binom{n}{r},$$

زیرا تعداد $\binom{n}{r}$ کلمه به فاصله r از $\mathbf{e}$ وجود دارد. این کران بالا مستقل از $\mathbf{e}$ است، پس اگر روی $\mathbf{e}$ میانگین بگیریم، آنگاه از (C.۵) نتیجه می‌شود

$$\overline{Pr(d(\mathbf{u}_j, \mathbf{u}_i + \mathbf{e}) \leq \rho)} \leq \frac{۱}{۲^n - ۱} \sum_{r \leq \rho} \binom{n}{r}.$$

این رابطه برای هر زوج $i \neq j$ برقرار است، پس اگر روی تمامی $M(M-1)$ زوج این چنینی عمل جمع انجام دهیم از (C.۴) نتیجه می‌شود

$$\begin{aligned}\overline{Pr_E} &< \frac{\delta}{2} + \frac{1}{M}.M(M-1).\frac{1}{2^n-1}\sum_{r\leq\rho}\binom{n}{r} \\ &\leq \frac{\delta}{2} + \frac{M}{2^n}\sum_{r\leq\rho}\binom{n}{r}. \qquad (C.6)\end{aligned}$$

حال فرض کنید η آن‌قدر کوچک انتخاب شده باشد که

$$Q+\eta < \frac{1}{2};$$

این به جهت $Q < \frac{1}{2}$ امکان‌پذیر است. در این‌صورت از تمرین ۷.۵ (با $\lambda = Q+\eta$، به‌طوری‌که $\rho = \lambda n$) نتیجه می‌شود

$$\sum_{r\leq\rho}\binom{n}{r} = \sum_{r\leq\lambda n}\binom{n}{r} \leq 2^{nH(\lambda)},$$

که H تابع آنتروپی دوتایی H_2 است، پس

$$\begin{aligned}\overline{Pr_E} &< \frac{\delta}{2} + 2^{-n}M.2^{nH(\lambda)} \\ &= \frac{\delta}{2} + M.2^{n(H(\lambda)-1)}. \qquad (C.7)\end{aligned}$$

نیاز به یک کران بالا روی جمله دوم در (C.۷) داریم. حال

$$\begin{aligned}\log_2(M.2^{n(H(\lambda)-1)}) &= \log_2 M + n(H(\lambda)-1) \\ &= n(R-1+H(\lambda)),\end{aligned}$$

که $R = \frac{1}{n}\log_2 M$ نرخ کد C است. فرض کنید ثابتی چون $\alpha > 0$ موجود باشد به‌قسمی‌که برای هر n داشته باشیم

$$R-1+H(\lambda) \leq \alpha;$$

در این‌صورت اگر $n \longrightarrow \infty$ آنگاه $n(R-1+H(\lambda)) \leq n\alpha \longrightarrow -\infty$، پس با افزایش n مقدار $M.2^{n(H(\lambda)-1)}$ به ۰ میل می‌کند. از این‌رو برای تمامی مقادیر به‌قدر کافی بزرگ n داریم

$$M.2^{n(H(\lambda)-1)} < \frac{\delta}{2},$$

و بنابراین از (C.۷) نتیجه می‌شود $\overline{Pr_e} < \delta$. چون میانگین Pr_E روی تمامی کدهای $\mathcal{C}$ کمتر از δ است، نتیجه می‌شود که حداقل یک کد $\mathcal{C}$ باید خاصیت $Pr_E < \delta$ را داشته باشد، و این چیزی است که در پی آن هستیم.

برای تکمیل اثبات، باید امکان انتخاب ثابت‌های مطرح شده را نشان دهیم. باید نشان دهیم برای اعداد مفروض $Q < \frac{1}{2}$ و $\varepsilon > 0$، می‌توان اعداد $\eta > 0$ و $\alpha < 0$ (مستقل از n) را پیدا کرد به‌قسمی‌که اگر n به‌قدر کافی بزرگ باشد آنگاه

(i) $Q + \eta < \frac{1}{2}$،

(ii) $C - \varepsilon \leq R < C$ (که $C = 1 - H(Q)$) و به‌ازای عددی چون $M \in \mathbf{N}$، $(R = \frac{1}{n}\log_2 M)$،

(iii) $R - 1 + H(Q + \eta) \leq \alpha$.

بدون از دست دادن کلیت مسئله، چون $C > 0$ می‌توانیم ε را آن‌قدر کوچک اختیار کنیم که $C - \varepsilon \geq 0$. حال $C = 1 - H(Q)$، تابع H پیوسته است، و $Q < \frac{1}{2}$، پس می‌توان $\eta > 0$ را آن‌قدر کوچک انتخاب کرد که

$$Q + \eta < \frac{1}{2},$$

که از آن (i) نتیجه شده و

$$1 - H(Q + \eta) \geq C - \frac{\varepsilon}{3}.$$

برای هر $n \geq 3/\varepsilon$ می‌توان عدد گویای $R = k/n$ $(k, n \in \mathbf{N})$ را به‌قسمی انتخاب کرد که

$$C - \varepsilon \leq R \leq C - \frac{2\varepsilon}{3}.$$

قرار می‌دهیم $M = 2^k$؛ پس $M \in \mathbf{N}$ و

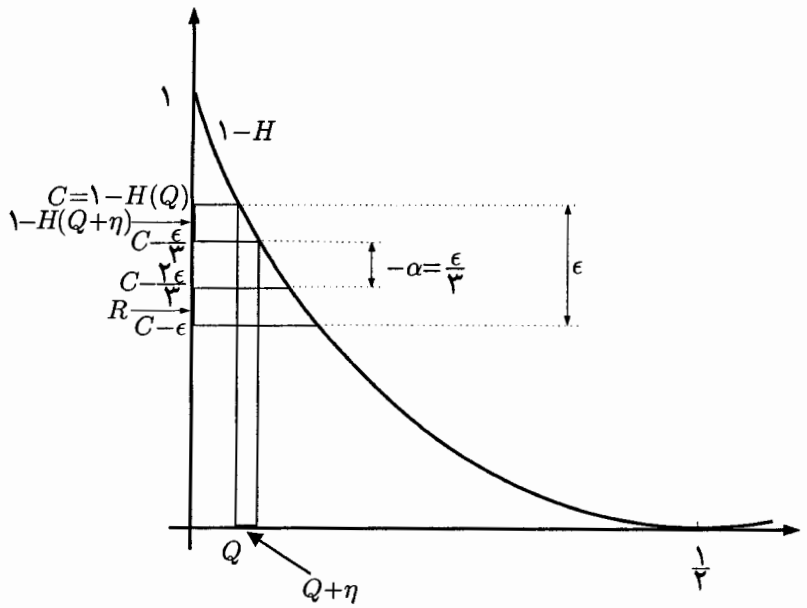
$$\frac{1}{n}\log_2 M = \frac{k}{n} = R,$$

و بنابراین (ii) برقرار می‌شود. چون $R \leq C - 2\varepsilon/3$ و $1 - H(Q + \eta) \geq C - \varepsilon/3$، پس

$$R - 1 + H(Q + \eta) \leq (C - \frac{2\varepsilon}{3}) - (C - \frac{\varepsilon}{3}) = -\frac{\varepsilon}{3};$$

این (iii) را با $\alpha = -\frac{\varepsilon}{3} < 0$ (مستقل از n) برقرار می‌سازد. ■

شکل (C.۱) رابطه بین کمیت‌های به‌کار رفته را توضیح می‌دهد.



شکل ۱.C

جواب تمرینات

۱.۱ از استقرا روی n استفاده کنید. اگر $n = ۰$ آنگاه $\mathcal{C}_n = \mathcal{C}$، پس $|w| \leq l$. اگر $n > ۰$ آنگاه $uw = v$ که v در $\mathcal{C}_{n-۱}$ یا $\mathcal{C}$ قرار دارد، پس $|w| \leq |v| \leq l$، به‌ترتیب بنابه فرض استقرا یا بنابر تعریف l. تنها تعداد $N = r + r^۲ + \cdots + r^l = r(r^l - ۱)/(r - ۱)$ کلمه r تایی غیرخالی w با خاصیت $|w| \leq l$ وجود دارد، پس برای هر n داریم $|\mathcal{C}_n| \leq N$. تنها تعداد ۲^N مجموعه متفاوت از کلمات این‌چنینی w وجود دارد، از این‌رو در بین مجموعه‌های $\mathcal{C}_۰$، ...، $\mathcal{C}_{۲^N}$ باید تکراری چون $\mathcal{C}_i = \mathcal{C}_j$ با $i < j \leq ۲^N$ موجود باشد. بنابر معادله (۳.۱)، هر $\mathcal{C}_n$ تنها به $\mathcal{C}$ و $\mathcal{C}_{n-۱}$ بستگی دارد، و بنابراین برای هر $k \geq ۰$ داریم $\mathcal{C}_{j+k} = \mathcal{C}_{i+k}$، پس هر $\mathcal{C}_n$ برابر $\mathcal{C}_۰$ یا $\mathcal{C}_۱$، یا ... یا $\mathcal{C}_{j-۱}$ است و بنابراین $\mathcal{C}_\infty = \mathcal{C}_۰ \cup \mathcal{C}_۱ \cup \cdots \cup \mathcal{C}_{j-۱}$. پس به‌محض مشاهده یک تکرار در میان مجموعه‌های متوالی $\mathcal{C}_۰$، $\mathcal{C}_۱$، ... مجموعه $\mathcal{C}_\infty$ ساخته می‌شود.

۲.۱ اگر $\mathcal{C} = \{۰۲, ۱۲, ۱۲۰, ۲۰, ۲۱\}$ آنگاه $\mathcal{C}_۱ = \{۰\}$، $\mathcal{C}_۲ = \{۲\}$، $\mathcal{C}_۳ = \{۰, ۱\}$، $\mathcal{C}_۴ = \{۲, ۲۰\}$، $\mathcal{C}_۵ = \{۰, ۱\}$؛ از تکرار $\mathcal{C}_۳ = \mathcal{C}_۵$ نتیجه می‌شود بسته به فرد یا زوج بودن $n \geq ۳$، $\mathcal{C}_n$ برابر $\{۰, ۱\}$ یا $\{۲, ۲۰\}$ می‌باشد، پس $\mathcal{C}_\infty = \mathcal{C}_۱ \cup \cdots \cup \mathcal{C}_۴ = \{۰, ۱, ۲, ۲۰\}$.
اگر $\mathcal{C} = \{۰۲, ۱۲, ۱۲۰, ۲۱\}$ آنگاه $\mathcal{C}_۱ = \{۰\}$، $\mathcal{C}_۲ = \{۲\}$، $\mathcal{C}_۳ = \{۱\}$، $\mathcal{C}_۴ = \{۲, ۲۰\}$، $\mathcal{C}_۵ = \{۱\}$؛ مجدداً از $\mathcal{C}_۳ = \mathcal{C}_۵$ نتیجه می‌شود بسته به فرد یا زوج بودن $n \geq ۳$، $\mathcal{C}_n$ برابر $\{۱\}$ یا $\{۲, ۲۰\}$ است و بنابراین $\mathcal{C}_\infty = \mathcal{C}_۱ \cup \cdots \cup \mathcal{C}_۴ = \{۰, ۱, ۲, ۲۰\}$.

۳.۱ اگر $\mathcal{C} = \{۰۲, ۱۲, ۱۲۰, ۲۰, ۲۱\}$ آنگاه از تمرین ۲.۱ نتیجه می‌شود $\mathcal{C}_\infty = \{۰, ۱, ۲, ۲۰\}$ که شامل کدکلمه ۲۰ می‌باشد، پس بنابر قضیه ۱۰.۱ کد $\mathcal{C}$ یکتا دکدپذیر نیست؛ به‌عنوان نمونه ۱۲۰۲۱۲۰ به دو فرم ۱۲۰.۲۱.۲۰ و ۱۲.۰۲.۱۲۰ دکد می‌شود. اگر $\mathcal{C} = \{۰۲, ۱۲, ۱۲۰, ۲۱\}$ آنگاه از تمرین ۲.۱ نتیجه می‌شود $\mathcal{C}_\infty = \{۰, ۱, ۲, ۲۰\}$ که جدا از $\mathcal{C}$ است و بنابراین $\mathcal{C}$ یکتا دکدپذیر است.

۴.۱ چون $u \in \mathcal{C}_1$، برای کلماتی چون $u', v' \in \mathcal{C}$ داریم $u'u = v'$، و بنابراین $\mathbf{t} = u'uw$ به‌فرم $u'v$ یا $v'w$ دکد می‌شود.

۵.۱ چون $\mathcal{C} \ni ۰۱,۰۱۲۱۲۰$، پس $\mathcal{C}_1 \ni ۲۱۲۰$؛ سپس از $\mathcal{C} \ni ۲۱۲$ نتیجه می‌شود $\mathcal{C}_2 \ni ۰$، و بنابراین $\mathcal{C} \ni ۰۱$ ایجاب می‌کند $\mathcal{C}_3 \ni ۱$؛ سپس از $\mathcal{C} \ni ۱۲۰$ نتیجه می‌شود $\mathcal{C}_4 \ni ۲۰$. بنابراین $\mathcal{C} \cap \mathcal{C}_\infty \ni ۲۰$.

۶.۱ چون $w \in \mathcal{C}_3$ پس عناصری مانند $u \in \mathcal{C}$ و $v \in \mathcal{C}_2$ وجوددارند به‌قسمی‌که یا (i): $uw = v$ و یا (ii): $vw = u$. چون $v \in \mathcal{C}_2$ پس عناصری مانند $u' \in \mathcal{C}$ و $v' \in \mathcal{C}_1$ وجود دارند که یا (a): $u'v = v'$ و یا (b): $v'v = u'$.

از $v' \in \mathcal{C}_1$ وجود اعضایی از $\mathcal{C}$ مانند u'' و v'' نتیجه می‌شود که $u''v' = v''$. حال $u, u', u'', v'', w \in \mathcal{C}$، از این‌رو در حالت‌های (a)(i)، (b)(i)، (a)(ii) و (a)(ii) به‌ترتیب مثال‌های یکتا دکدناپذیر زیر را داریم:

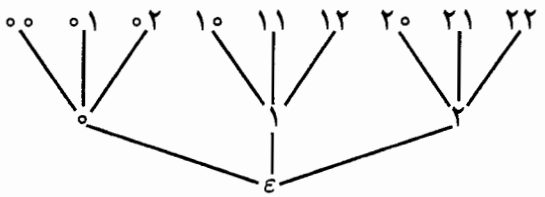
$$\begin{aligned} u''u'uw &= u''u'v = u''v' = v'', \\ v''uw &= u''v'v = u''u', \\ u''u'u &= u''u'vw = u''v'w = v''w, \\ v''u &= u''v'vw = u''u'w. \end{aligned}$$

۷.۱ برای هر $n \geq ۱$، و هر یک از دو کد داده شده، هر کد $\mathcal{C}_n$ غیرخالی است، و بنابراین تمامی کددنباله‌های نامتناهی به‌صورت یکتا دکد نمی‌شوند. به‌عنوان مثال ۱۲۰۲۱۲۱۲۱... به‌صورت‌های ۱۲۰.۲۱.۲۱... یا ۱۲.۰۲.۱۲.۱۲... دکد می‌شود.

۸.۱ $\mathcal{C}_1 = \{۱, ۱۱\}$ و $\mathcal{C}_2 = \{۱, ۱۱\}$، پس برای هر $n \geq ۱$ داریم $\mathcal{C}_n = \{۱, ۱۱\}$؛ بنابراین $\mathcal{C}_\infty = \{۱, ۱۱\}$، که مجزا از $\mathcal{C}$ است و بنابراین بنابر قضیه ۱۰.۱ کد $\mathcal{C}$ یکتا دکدپذیر است. تا به پایان رسیدن دنباله ۱ ها صبر کنید؛ اگر تعداد ۱ ها برابر k است، که $k \equiv ۰, ۱, ۲ \pmod{۳}$، آنگاه به‌صورت یکتا به فرم‌های $۰.(۱۱۱)^{k/۳}$، $۰۱.(۱۱۱)^{(k-۱)/۳}$ و یا $۰۱۱.(۱۱۱)^{(k-۲)/۳}$ دکد کنید.

۹.۱ بله. یک سمبل آغازین ۰ اشاره به w_1 دارد، درحالی‌که یک ۱ اشاره به شروع w_2، w_3، یا w_4 دارد؛ در حالت دوم یک سمبل دوم ۰ اشاره به w_2 دارد درحالی‌که یک ۱ اشاره به w_3 یا w_4 دارد؛ در این حالت دوم، یک سمبل ۰ به w_3 و یک سمبل ۱ به w_4 مربوط می‌شود.

۱۰.۱ تا لایه دوم نمودار زیر را داریم؛



حال به هر یک از رئوس v در لایه دوم سه رأس v۰، v۱، v۲ را ضمیمه کنید.

۱۱.۱ $C = \{۰, ۱۰, ۱۱۰, ۱۱۱, ۲۰۰۰\}$ یک مثال است. خیر، زیرا انتخاب‌های متوالی منجر به حذف درصدهای $\frac{۱}{۳}$،$\frac{۱}{۳}$،$\frac{۱}{۴}$،$\frac{۱}{۴}$،$\frac{۱}{۴}$،$\frac{۱}{۴}$ از T^* شده و $\frac{۱۰}{۴} > ۱$ تناقض است.

۱۲.۱ براساس نامساوی کرفت جواب منفی است، زیرا $\sum r^{-l_i} = \frac{۲۸}{۲۷} > ۱$. برای یک مثال مجموعه $\{۰, ۱۰, ۱۱, ۱۲, ۲۰, ۲۱, ۲۲۰, ۲۲۱, ۲۲۲\}$ را درنظر بگیرید. سه انتخاب برای کدکلمه‌های به‌طول ۱ وجود دارد (۰، ۱، یا ۲)، و سپس $\binom{۶}{۵} = ۶$ انتخاب برای ۵ کدکلمه به‌طول ۲، و بنابراین انتخاب‌های یکتایی برای سه کدکلمه به‌طول ۳؛ پس تعداد یک چنین کدهایی برابر $۱۸ = ۱ \times ۶ \times ۳$ است.

۱۳.۱ اگر $j \geq ۲$ آنگاه $\mathbf{t} = \mathbf{t}'w$ که $w \in \{۰, ۱۰, ۱۱\}$. اگر $w = ۰$ آنگاه $N_{j-۱}$ امکان برای $\mathbf{t}'$ (به‌طول $j - ۱$) وجود دارد؛ اگر $w = ۱۰$ یا $w = ۱۱$ آنگاه در هر حالت تعداد $N_{j-۲}$ امکان برای $\mathbf{t}'$ (به‌طول $j - ۲$) وجود دارد. از این‌رو $N_j = N_{j-۱} + ۲N_{j-۲}$. این رابطه خطی بازگشتی درجه دوم دارای معادله معین $\lambda^۲ = \lambda + ۲$ با ریشه‌های $\lambda = ۲, -۱$ است و بنابراین جواب عمومی آن به‌فرم $N_j = A.۲^j + B.(-۱)^j$ است. از شرایط اولیه $N_۱ = ۱$ $(\mathbf{t} = ۰)$ و $N_۲ = ۳$ $(\mathbf{t} = ۰۰, ۱۰, ۱۱)$ نتیجه می‌شود $A = \frac{۲}{۳}$ و $B = \frac{۱}{۳}$، پس $N_j = (۲^{j+۱} + (-۱)^j)/۳$. (برای روابط بازگشتی به فصل ۴ از [An74] مراجعه کنید.)

۱۴.۱ در اثبات قضیه ۲۰.۱، تعداد $r^{l_۱}$ انتخاب برای $w_۱$، و سپس برای $w_۲$ $r^{l_۲} - r^{l_۲-l_۱} - r^{l_۲-l_۲} = r^{l_۲}(۱ - ۲^{-l_۱} - r^{-l_۲}) = r^{l_۲} - r^{l_۲-l_۱}$ انتخاب، و بعداً $w_۳$ انتخاب برای وجود داشته و به همین روش تعداد انتخاب‌های بعدی مشخص می‌شود. پس تعداد انتخاب‌های ممکن برای $w_۱$،...،w_q برابر است با

$$r^{l_۱+l_۲+\cdots+l_q}(۱ - r^{-l_۱}) \cdots (۱ - r^{-l_۱} - \cdots - r^{-l_{q-۱}}).$$

۱۵.۱ C جامع است اگر و فقط اگر هر برگ از $T^{\leq l}$ بالای یک کدکلمه باشد. کدهای مثال‌های ۱۶.۱ و ۱۸.۱ جامع هستند.

۱۶.۱ از اثبات قضیه ۲۰.۱ پیروی کنید: C جامع است اگر و فقط اگر هر برگ از $T^{\leq l}$ بالای یک کدکلمه باشد؛ تعداد برگ‌ها r^l است و هر کدکلمه به‌طول l_i زیر r^{l-l_i} برگ قرار

دارد و بنابراین $r^l \leq \sum_i r^{l-l_i}$، یعنی $\sum_i r^{-l_i} \geq 1$. تساوی برقرار خواهد بود اگر و فقط اگر هر برگ دقیقاً بالای یک کدکلمه باشد، یعنی این‌که C پیشوند است.

۱۷.۱ بنابر تمرین ۱۶.۱، اگر (ب) درست باشد آنگاه (الف) معادل (ج) خواهد بود؛ پس از (الف) و (ب) نتیجه می‌شود (ج)، و از (ب) و (ج) نتیجه می‌شود (الف). اگر (الف) و (ج) درست باشند، آنگاه در اثبات قضیه ۲۰.۱ هر برگ از $T^{\leq l}$ بالای یک کدکلمه قرار دارد و بنابراین (ب) نتیجه می‌شود. اگر $T = \mathbf{Z}_2$ آنگاه کدهای $\{0\}$، $\{0, 1, 00\}$ و $\{0, 00, 01\}$ به‌ترتیب تنها در (الف)، (ب)، و (ج) صدق می‌کنند و بنابراین هیچ‌یک از این سه خاصیت لزوماً از یکی از دو خاصیت دیگر نتیجه نمی شود.

۱.۲ فرض کنید $p_i > p_j$ و $l_i > l_j$. با تعویض کدکلمه‌های w_i و w_j در C یک کد لحظه‌ای C^* به‌دست می آید. جملات $p_i l_i$ و $p_j l_j$ در $L(C)$ با جملات $p_i l_j$ و $p_j l_i$ در $L(C^*)$ تعویض می‌شوند. در این‌صورت

$$(p_i l_i + p_j l_j) - (p_i l_j + p_j l_i) = (p_i - p_j)(l_i - l_j) > 0$$

و بنابراین $L(C) > L(C^*)$، که در تناقض با بهینه بودن C می‌باشد. از این‌رو $l_i \leq l_j$.

۲.۲ از $\mathcal{S}$ یک بردار $\mathbf{p} = (p_1, \ldots, p_q) \in \mathbf{R}^q$ با خواص $p_i \geq 0$ و $\sum p_i = 1$، و از یک کد C یک بردار $\mathbf{l} = (l_1, \ldots, l_q) \in \mathbf{N}^q \subset \mathbf{R}^q$ نتیجه می‌شود به‌طوری‌که $L(C) = \sum p_i l_i = \mathbf{p.l}$. برای یک $\mathbf{p}$ مفروض، مسئله نشان دادن این است که یک کد لحظه‌ای وجود دارد که برای آن $\mathbf{p.l}$ می‌نیمم ممکن است. اثبات قضیه ۳.۲ نشان می‌دهد که چون $l_i \in \mathbf{N}$، پس تنها تعداد متناهی مقدار ممکن برای $\mathbf{p.l}$ وجود دارد. در بین تعداد متناهی امکان موجود متناظر با کدهای لحظه‌ای می‌توان کمترین مقدار را انتخاب کرد که متناظر با یک کد بهینه می‌شود.

۳.۲ یک جواب $C = \{0, 10, 1100, 1101, 1110, 1111\}$ با میانگین $\sum p_i l_i = 2.2$ است. یک امکان دیگر $C = \{1, 00, 011, 0100, 01010, 01011\}$ است، بنابراین C و $\{l_i\}$ یکتا نیستند. با این حال $L(C)$ بنابر بهینه بودن کدهای هافمن یکتا است.

۴.۲ برای $q = 3$، کد $C = \{0, 10, 11\}$ میانگین طول $L(C) = p_1 + 2p_2 + 2p_3 = 2 - p_1$ دارد. وقتی $q = 4$، داریم $l_i = (1, 2, 3, 3)$ یا $l_i = (2, 2, 2, 2)$ زیرا $p_3 + p_4 \leq p_1$ یا $p_3 + p_4 \geq p_1$، در نتیجه به‌ترتیب $L(C) = p_1 + 2p_2 + 3p_3 + 3p_4 = 3 - 2p_1 - p_2$ یا $L(C) = 2p_1 + 2p_2 + 2p_3 + 2p_4 = 2$.

۵.۲ در تمرین ۳.۲، احتمال‌های ادغام شده p'، p''، ... عبارت هستند از ۰.۱، ۰.۲، ۰.۳، ۰.۶، ۱ با مجموع ۲.۲. در تمرین ۴.۲ با $q = 3$ این‌ها عبارت‌اند از $p' = p_2 + p_3$ و $p'' = 1 - p_1 = 2 + 1 + p_3 + p_2 = p' + p''$؛ برای $q = 4$ داریم $p' = p_3 + p_4$، $p'' = p_2 + p_3 + p_4$ یا $p'' = p_1 + p_2$ بسته به این‌که $p_3 + p_4 \leq p_1$ یا $p_3 + p_4 \geq p_1$، و $p''' = 1$. پس به‌ترتیب $p' + p'' + p''' = 1 + p_2 + 2p_3 + 2p_4 = 3 - 2p_1 - p_2$ یا $p' + p'' + p''' = 2$.

۶.۲ اثبات این‌که کدهای هافمن بهینه هستند از مقایسه کدهای هافمن و کدهای بهینه نتیجه می‌شود. این فرض می‌کند که هر منبعی یک کد بهینه دارد، و بنابراین بحث دوری می‌شود.

۷.۲ برای حالت دوتایی، $\mathcal{C} = \{00, 10, 010, 110, 111, 0110, 01110, 01111\}$ با $L(\mathcal{C}) = 2.72$. برای حالت سه‌تایی، $\mathcal{C} = \{0, 10, 11, 12, 20, 21, 220, 221\}$ با $L(\mathcal{C}) = 1.77$.

۸.۲ بحث شبیه به بخش ۳.۲ است به‌جز این‌که سمبل‌های s_{q-r+1}، ... s_q آمیخته می‌شوند و

$$\begin{aligned} L(\mathcal{C}) - L(\mathcal{C}') &= p_{q-r+1}(l+1) + \cdots + p_q(l+1) - (p_{q-r+1} + \cdots + p_q)l \\ &= p_{q-r+1} + \cdots + p_q = p'. \end{aligned}$$

سمبل‌های اضافه شده دیگر تأثیری روی $L(\mathcal{C})$ ندارند زیرا احتمال آنها صفر است.

۹.۲ $\mathcal{S}^3$ دارای $8 = 2^3$ سمبل با احتمال $\frac{8}{27}$، $\frac{4}{27}$، $\frac{4}{27}$، $\frac{4}{27}$، $\frac{2}{27}$، $\frac{2}{27}$، $\frac{2}{27}$، $\frac{1}{27}$ است. در کدگذاری هافمن احتمال‌های ادغام‌شده عبارت‌اند از $\frac{3}{27}$، $\frac{4}{27}$، $\frac{7}{27}$، $\frac{8}{27}$، $\frac{11}{27}$، $\frac{16}{27}$، ۱، با مجموع $L_3 = \frac{76}{27}$.

۱۰.۲ تعداد ۲۴ کد دوتایی بهینه وجود دارد، تعداد !۴ جایگشت مجموعه $\{00, 01, 10, 11\}$ (این کدها دارای $L(\mathcal{C}) = 2$ هستند، درحالی‌که هر کد لحظه‌ای دیگری دارای $L(\mathcal{C}) \geq 2.1$ است). در بین اینها هشت کد هافمن وجود دارد، آنهایی که در آنها دو کدکلمه آخر (دو کلمه با کمترین احتمال) هم‌زاد هستند: در ساخت کدهای $\mathcal{C}''$، $\mathcal{C}'$ و $\mathcal{C}$ دو امکان در هر مرحله وجود دارد ($w'0$، $w'1$ یا $w'1$، $w'0$) و بنابراین $8 = 2^3$ امکان برای $\mathcal{C}$ وجود دارد.

۱۱.۲ از نامساوی‌های داده شده نتیجه می‌شود $s' = s_{q-1} \vee s_q$، و سپس با استقرا روی k نتیجه می‌شود برای هر $1 < k \leq q-1$ داریم $s^{(k-1)} = s_{q-k} \vee \cdots \vee s_q$. پس برای $i \leq q-1$، s_i به تعداد i دفعه آمیخته شده (در $\mathcal{S}^{(q-1-i)}$،...،$\mathcal{S}^{(q-2)}$) و بنابراین $l_i = i$، درحالی‌که s_q تعداد $q-1$ بار آمیخته شده و از این‌رو $l_q = q-1$. در هر یک از این $q-1$ مرحله تنها دو انتخاب در تخصیص کدکلمه‌ها وجود دارد: از کدکلمه $w^{(k)}$ برای $s^{(k)}$ دو کدکلمه $w^{(k)}0$ و $w^{(k)}1$ برای s_{q-k} و $s^{(k-1)}$ نتیجه می‌شود. از این‌رو تعداد 2^{q-1} کد هافمن دوتایی برای S وجود دارد. احتمال‌های $p_i = 2^{-i}$، $1 \leq i \leq q-1$، و $p_q = 2^{1-q}$ در شرایط داده شده صدق می‌کنند، زیرا برای $1 \leq i \leq q-3$ داریم $p_i > 2^{-i-1} = p_{i+2} + \cdots + p_q$.

۱۲.۲ در کدگذاری r تایی هافمن، یک کدکلمه به‌طول l چون $w' \in C'$ با r کدکلمه به‌طول $l+1$ در C جایگزین می‌شود، پس $\sigma(C) - \sigma(C') = r(l+1) - l = (r-1)l + r$. چون r ثابت است می‌توان $\sigma(C)$ را با می‌نیمم کردن l در هر مرحله از الگوریتم می‌نیمم کرد. برای این منظور سمبل آمیخته s' باید قبل از تمامی سمبل‌های هم‌احتمال با خود در S' قرار گیرد. در مثال داده شده، $s' = s_3 \vee s_4$ دارای احتمال $\frac{1}{3}$ است و در عین حال $p_1 = p_2 = \frac{1}{3}$. پس سه محل ممکن برای s' در S' قرار دارد؛ با قرار دادن آن قبل از p_1 مطمئن می‌شویم که $l = 1$ خواهد بود و نه ۲، از این‌جا مقدار $\sigma(C) = 2+2+2+2 = 8$ به‌جای $\sigma(C) = 1+2+3+3 = 9$ نتیجه می‌شود.

۱۳.۲ با به‌کار بردن احتمال‌های p_i داده شده برای سمبل‌های s_i، یک کد هافمن دوتایی $C = \{w_1, \ldots, w_q\}$ بسازید. برای هر k، فرض کنید T_k مجموعه سمبل‌هایی باشد که کدکلمه آنها در موقعیت k برابر ۱ است، و فرض کنید Q_k این سؤال باشد ' آیا s در T_k است؟ '. در این‌صورت با درخواست $Q_1, Q_2, \ldots$ کدکلمه w_i و بنابراین s_i بعد از $l_i = |w_i|$ سؤال مشخص می‌شود، بنابراین میانگین تعداد سؤالات لازم برابر $L(C) = \sum_i p_i l_i$ است.

مشابهاً با نسبت دادن سمبل‌های ۱ و ۰ به‌ترتیب به بله و خیر، هر دنباله دیگر از سوالات متناظر با کد پیشوند دوتایی دیگری برای S می‌شود، بنابراین چون کد هافمن بهینه است متناظر با بهترین دنباله ممکن می‌شود.

۱۴.۲ بله: اگر $q = 3$ آنگاه $L(C) = 5/3$؛ S^2 دارای ۹ سمبل هم‌احتمال است و بنابراین بنابر (۴.۲) خواهیم داشت $L_2 = 29/9$، پس $L_2/2 = 29/18 < L(C)$. اگر به‌ازای عددی چون l داشته باشیم $q = 2^l$ آنگاه در C هر l_i برابر l بوده و از این‌رو $L(C) = l$؛ مشابهاً S^n دارای 2^{ln} سمبل هم‌احتمال است، پس $L_n = l$ و از این‌رو $L_n/n = L(C)$.

۱.۳ $H_۲(\mathcal{S}) = \sum_i p_i \log_۲(۱/p_i) \approx ۲.۶۸۱$ و $H_۳(\mathcal{S}) = \sum_i p_i \log_۳(۱/p_i) \approx ۱.۶۹۱$.
کدهای هافمن دوتایی و سه‌تایی به‌ترتیب دارای میانگین طول $L(\mathcal{C}) = ۲.۹۲$ و $L(\mathcal{C}) = ۱.۷۷$ هستند.

۲.۳ فرض کنید $\mathcal{S}$ دارای احتمال‌های $P = (۲^{-۱}, ۲^{-۲}, ۲^{-۳}, \ldots, ۲^{۲-q}, ۲^{۱-q}, ۲^{۱-q})$ باشد؛ همچنین کد $\mathcal{C} = \{۰, ۱۰, ۱۱۰, \ldots, ۱\ldots۱۰, ۱\ldots۱۱\}$ با طول دنباله $l = (۱, ۲, ۳, \ldots, q-۲, q-۱, q-۱)$ را درنظر بگیرید. در این‌صورت:

$$H_۲(\mathcal{S}) = ۲^{-۱}.۱ + ۲^{-۲}.۲ + ۲^{-۳}.۳ + \cdots + ۲^{۲-q}.(q-۲) + ۲.۲^{۱-q}.(q-۱) = L(\mathcal{C}).$$

۳.۳ $H_۲(\mathcal{S}) = \sum_i p_i \log_۲(۱/p_i) \approx ۲.۱۴۴$. بنابر تمرین ۳.۲ یک کد هافمن $\mathcal{C}$ دارای $L(\mathcal{C}) = ۲.۲$ بوده، و از این‌رو $\eta = H/L \approx ۲.۱۴۴/۲.۲ \approx ۰.۹۷۵$.

۴.۳ یک کد شانون — فانو دارای $l_i = \lceil \log_۲(۱/p_i) \rceil = ۲, ۲, ۴, ۴, , ۵, ۵$ است، پس $L = ۲.۷$ و $\eta \approx ۲.۱۴۴/۲.۷ \approx ۰.۷۹۴$.

۵.۳ $H_۲(\mathcal{S}) = -\frac{۴}{۵}\log_۲ \frac{۴}{۵} - \frac{۱}{۵}\log_۲ \frac{۱}{۵} = \log_۲ ۵ - \frac{۸}{۵}$. در $\mathcal{S}^n$ برای هر $۰ \le k \le n$ تعداد $\binom{n}{k}$ سمبل با احتمال $(۴/۵)^k(۱/۵)^{n-k} = ۴^k/۵^n$ وجود دارد، هر یک از آنها با کدکلمه به‌طول $\lceil \log_۲(۵^n/۴^k) \rceil = \lceil n\log_۲ ۵ \rceil - ۲k$ نمایش داده می‌شود، پس اگر a_n معرف $\lceil n \log_۲ ۵ \rceil$ باشد آنگاه، مانند بخش ۷.۳، یک کد دوتایی شانون — فانو برای $\mathcal{S}^n$ دارای میانگین طول L_n است:

$$L_n = \sum_{k=۰}^{n} \binom{n}{k} \frac{۴^k}{۵^n}(a_n - ۲k) = \frac{۱}{۵^n}\left(a_n \sum_{k=۰}^{n}\binom{n}{k}۴^k - ۲\sum_{k=۰}^{n} k\binom{n}{k}۴^k\right) = a_n - \frac{۸n}{۵}.$$

با افزایش n داریم $a_n/n \longrightarrow \log_۲ ۵$، و بنابراین $\frac{۱}{n}L_n \longrightarrow \log_۲ ۵ - \frac{۸}{۵} = H_۲(\mathcal{S})$.

۶.۳ $\mathcal{S}^n$ دارای q^n سمبل است که همگی دارای احتمال $۱/q^n$ می‌باشند، از این‌رو هر کدام با کدکلمه‌ای r تایی به‌طول $\lceil \log_r q^n \rceil = \lceil n \log_r q \rceil$ نمایش داده می‌شوند. پس $L_n = \lceil n \log_r q \rceil$ و از این‌رو $\frac{۱}{n}L_n = \lceil n\log_r q \rceil / n$ که با افزایش n به $\log_r q = H_r(\mathcal{S})$ میل می‌کند.

۷.۳ تعریف کنید $g(x) = f(e^{-x})$، یک تابع اکیداً صعودی روی $[۰, +\infty)$ با خاصیت $g(x+y) = g(x) + g(y)$. با قرار دادن $x = y = ۰$ ملاحظه می‌شود $g(۰) = ۰$.
تعریف کنید $c = g(۱)$، پس $g(۰) = ۰ < c$. نشان می‌دهیم که $g(x) = cx$، پس $f(x) = -c\ln x = -\log_r x$. بنویسید $r = e^{۱/c} > ۰$. با استقرا روی n ملاحظه

می‌شود که برای هر $n \geq ۰$ داریم $g(۲^n) = c۲^n$. نیز $c = g(۱) = g(\frac{۱}{۲}) + g(\frac{۱}{۲})$، پس $g(\frac{۱}{۲}) = c/۲$، و استقرا روی n نشان می‌دهد $g(۲^{-n}) = c۲^{-n}$. هر $x \geq ۰$ دارای یک نمایش دوتایی $x = \sum_{n=-\infty}^{N} a_n ۲^n$ است که $a_n \in \{۰, ۱\}$، پس برای هر $M \leq N$ داریم $\sum_{n=M}^{N} a_n ۲^n \leq x \leq \sum_{n=M}^{N} a_n ۲^N + ۲^M$. از به‌کار بردن g روی این نامساوی‌ها و استفاده از خواص جمعی و صعودی بودن آن نتیجه می‌شود

$$\sum_{n=M}^{N} ca_n ۲^n \leq g(x) \leq \sum_{n=M}^{N} ca_n ۲^n + c۲^M.$$

با تقسیم بر c و میل دادن M به $-\infty$ ملاحظه می‌شود $g(x)/c = x$.

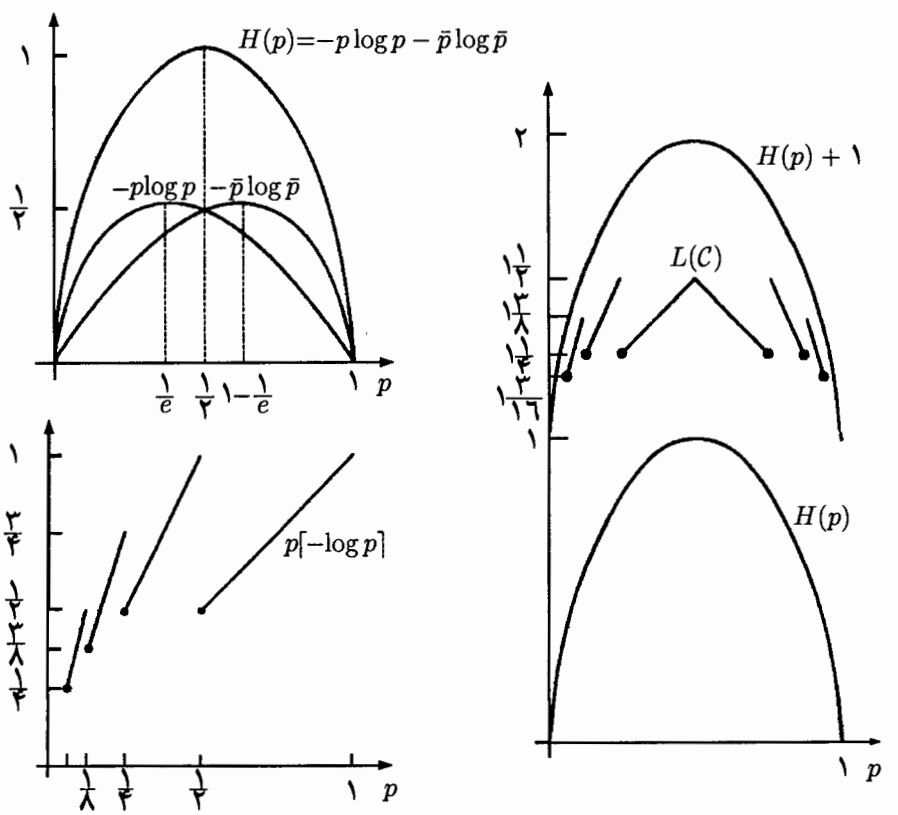
۸.۳ $\{s_i \in \{۲, ۳, \ldots, ۱۲\}$ و $P = \frac{۱}{۳۶}(۱, ۲, ۳, ۴, ۵, ۶, ۵, ۴, ۳, ۲, ۱)$، بنابراین

$$H_۲(\mathcal{S}) = \sum_i p_i \log_۲(۱/p_i) \approx ۳.۲۷۴۵.$$

در کدگذاری هافمن احتمال‌های متوالی ادغام شده p' عبارت هستند از $\frac{۲}{۳۶}$، $\frac{۴}{۳۶}$، $\frac{۵}{۳۶}$، $\frac{۷}{۳۶}$، $\frac{۸}{۳۶}$، $\frac{۱۰}{۳۶}$، $\frac{۱۱}{۳۶}$، $\frac{۱۵}{۳۶}$، $\frac{۲۱}{۳۶}$، ۱ با مجموع $L(C) = ۱۱۹/۳۶ \approx ۳.۳۰۵۵۵$.

در کدگذاری شانون — فانو داریم $l_i = \lceil \log_۲(۱/p_i) \rceil = ۶, ۵, ۴, ۴, ۳, ۳, ۳, ۴, ۴, ۵, ۶$، پس $L(C) = \sum_i p_i l_i = ۱۳۶/۳۶ \approx ۳.۷۷۷$.

۹.۳



۱۰.۳ بنابر نتیجه ۱۲.۳، $L(C) = H_r(S)$ اگر و فقط اگر S دارای احتمال‌های $p_i = r^{e_i}$ باشد که e_i عدد صحیح نامثبت است. در این حالت

$$\sum_{i=1}^{q} r^{e_i} = \sum_{i=1}^{q} p_i = 1,$$

پس اگر $e = \min e_i$ آنگاه $\sum_{i=1}^{q} r^{e_i - e} = r^{-e}$ که $e_i - e, -e \geq 0$؛ چون $r^{e_i - e}, r^{-e} \equiv 1 \bmod (r-1)$، پس $q \equiv 1 \bmod (r-1)$. بعکس اگر $q = 1 + k(r-1)$، آنگاه برای هر $1 \leq l \leq k-1$ تعداد $r-1$ سمبل هریک با احتمال r^{-l} و تعداد r سمبل هریک با احتمال r^{-k} درنظر بگیرید، پس

$$\sum_{i=1}^{q} p_i = (r-1)\sum_{l=1}^{k-1} r^{-l} + r.r^{-k} = 1.$$

سپس از نتیجه ۱۲.۳ ملاحظه می‌شود $L(C) = H_r(S)$.

۱۱.۳ $H_2(S) = -\frac{3}{4}\log_2 \frac{3}{4} - \frac{1}{4}\log_2 \frac{1}{4} = \log_2 4 - \frac{3}{4}$. در S^n برای هر $0 \leq k \leq n$ تعداد $\binom{n}{k}$ سمبل با احتمال $(3/4)^k(1/4)^{n-k} = 3^k/4^n$ وجود دارد که به هریک از آنها کدکلمه‌ای به‌طول $\lceil \log_2(4^n/3^k) \rceil = \lceil n\log_2 4 \rceil - k$ نسبت داده می‌شود، اگر $a_n = \lceil n \log_2 4 \rceil$ آنگاه یک کد شانون — فانو سه‌تایی برای S^n دارای میانگین طول L_n است:

$$L_n = \sum_{k=0}^{n} \binom{n}{k} \frac{3^k}{4^n}(a_n - k) = \frac{1}{4^n}\left(a_n \sum_{k=0}^{n} \binom{n}{k} 3^k - \sum_{k=0}^{n} k \binom{n}{k} 3^k\right) = a_n - \frac{3n}{4}.$$

با افزایش n مقدار a_n/n به $\log_2 4$، و بنابراین $\frac{1}{n}L_n$ به $H_2(S) = \log_2 4 - \frac{3}{4}$ میل می‌کند. در کدگذاری دوتایی شانون — فانو، به یک سمبل با احتمال $3^k/4^n$ یک کدکلمه به‌طول $\lceil \log_2(4^n/3^k) \rceil = 2n - b_k$ نسبت داده می‌شود که $b_k = \lfloor k \log_2 3 \rfloor$، پس

$$L_n = \sum_{k=0}^{n} \binom{n}{k} \frac{3^k}{4^n}(2n - b_k) = 2n - \frac{1}{4^n}\sum_{k=0}^{n} \binom{n}{k} 3^k b_k.$$

هیچ روش ساده‌ای برای تعیین مقدار این مجموع آخر وجود ندارد؛ با این حال از $k\log_2 3 - 1 < b_k \leq k \log_2 3$ نتیجه می‌شود

$$\frac{1}{4^n}\sum_{k=0}^{n} \binom{n}{k} 3^k (k\log_2 3 - 1) < \frac{1}{4^n}\sum_{k=0}^{n} \binom{n}{k} 3^k b_k \leq \frac{1}{4^n}\sum_{k=0}^{n} \binom{n}{k} 3^k k \log_2 3$$

و از این‌رو

$$\frac{3n}{4}\log_2 3 - 1 < \frac{1}{4^n}\sum_{k=0}^{n} \binom{n}{k} 3^k b_k \leq \frac{3n}{4}\log_2 3,$$

پس با افزایش n مقدار $\frac{1}{n}L_n$ به $H_2(S) = \frac{3}{4}\log_2 3 - 2$ میل می‌کند.

۱۲.۳ فرض کنید $p_1 = 1 - \delta$ و $p_2 = \cdots = p_q = \delta/(q-1)$، که $0 < \delta < 1$. در این‌صورت با میل کردن δ به ۰ مقدار $H_r(S) = -(1-\delta)\log_r(1-\delta) - \delta\log_r(\delta/(q-1))$ به ۰ می‌کند (زیرا $x\log_r x$ به ۰ میل می‌کند هرگاه x به ۰ یا ۱ میل کند)، پس برای مقادیر به‌قدر کافی کوچک از δ خواهیم داشت $H_r(S) < \varepsilon$. هر کد لحظه‌ای $\mathcal{C}$ دارای $L(\mathcal{C}) \geq 1$ است، پس $L(\mathcal{C}) > 1 + H_r(S) - \varepsilon$.

۱۳.۳ تعریف کنید $H_r(S) = \sum_{k=1}^{\infty} p_k \log_r(1/p_k) = -\sum_{k=1}^{\infty} p_k \log_r p_k$. اگر $p_k = 2^{-k}$ آنگاه $H_2(S) = \sum_{k=1}^{\infty} 2^{-k}\log_2 2^k = \sum_{k=1}^{\infty} 2^{-k}k = 2$. برای مرحله آخر از $(1-x)^{-1} = 1 + x + x^2 + \cdots$ مشتق گرفته، در x ضرب کنید و سپس قرار دهید $x = \frac{1}{2}$. کد پیشوند $\mathcal{C} = \{0, 10, 110, 1110, \ldots\}$ لحظه‌ای است و

$$L(\mathcal{C}) = \sum_{k=1}^{\infty} 2^{-k}k = 2 = H_2(S).$$

۱۴.۳ اگر $X_n = s_i$، آنگاه ابهام راجع به X_{n+1} برابر آنتروپی شرطی $H(S \mid X_n = s_i) = -\sum_j p_{ij}\log p_{ij}$ است؛ از متوسط‌گیری روی s_i نتیجه می‌شود $-\sum_i \sum_j p_i p_{ij}\log p_{ij} = \sum_i p_i e\left(-\sum_j p_{ij}\log p_{ij}\right)$ که میانگین ابهام روی S است. اعداد $p_i p_{ij} = Pr(X_n = s_i, X_{n+1} = s_j)$ (برای S^2) تشکیل یک توزیع احتمال می‌دهند همچنان‌که اعداد $p_i p_j$ تشکیل یک توزیع احتمال برای T^2 می‌دهند، براساس نتیجه ۹.۳ داریم $-\sum_i \sum_j p_i p_{ij}\log(p_i p_{ij}) \leq -\sum_i \sum_j p_i p_{ij}\log(p_i p_j)$ و از این‌رو (به دلیل خاصیت جمعی لگاریتم)

$$-\sum_i \sum_j p_i p_{ij}\log p_{ij} \leq -\sum_i \sum_j p_i p_{ij}\log p_j.$$

چون $\sum_i p_i p_{ij} = p_j$، از این نتیجه می‌شود $H(S) \leq H(T)$. از نتیجه ۹.۳ تساوی ایجاد می‌شود اگر و فقط اگر برای هر i و j داشته باشیم $p_i p_{ij} = p_i p_j$، یعنی X_{n+1} و X_n مستقل باشند. این‌گونه تفسیر می‌شود که دانستن احتمال‌های p_{ij} معمولاً ابهام ما را راجع به S کاهش می‌دهد. چون $\sum_i p_i p_{ij} = p_j$، (p_i) یک بردار ویژه ماتریس (p_{ij}) با مقدار ویژه $\lambda = 1$ است که در $\sum p_i = 1$ صدق می‌کند. در این حالت $p_1 = p_3 = \frac{1}{4}$ و $p_2 = \frac{1}{2}$، پس $H(T) = 3/2$ و $H(S) = (2 + 9\log 3)/12 \approx 1.355$.

۱.۴ اگر Γ دارای سمبل‌های ورودی و خروجی به‌ترتیب a_i و b_j بوده و اگر Γ' دارای سمبل‌های ورودی و خروجی به‌ترتیب b_j و c_k باشد آنگاه

$$Pr(c_k|a_i) = \sum_j Pr(b_j|a_i)Pr(c_k|b_j).$$

این قاعده ضرب ماتریس‌ها است، پس اگر M و M' به‌ترتیب ماتریس کانال‌های Γ و Γ' باشند آنگاه کانال ترکیبی $\Gamma' \circ \Gamma$ دارای ماتریس کانال MM' خواهد بود. در حالت کلی‌تر، اگر کانال‌های $\Gamma_1, \ldots, \Gamma_n$ به‌ترتیب دارای ماتریس کانال $M_1, \ldots, M_n$ باشند و خروجی کانال Γ_i ورودی کانال Γ_{i+1} باشد آنگاه استقرا روی n نشان می‌دهد که $M_1 \ldots M_n$ ماتریس کانال $\Gamma_1 \circ \cdots \circ \Gamma_n$ است.

۲.۴

(الف) $Q_{00} = pP/q$ و $Q_{10} = \overline{p}\,\overline{P}/q$، پس $Q_{10} < Q_{00}$ اگر و فقط اگر $pP < \overline{p}\,\overline{P}$؛ مشابهاً $Q_{01} = p\overline{P}/\overline{q}$ و $Q_{11} = \overline{p}P/\overline{q}$، پس $Q_{11} < Q_{01}$ اگر و فقط اگر $p\overline{P} < \overline{p}P$. معادلاً، $p = p(P+\overline{P}) < (p+\overline{p})P = P$ و $p = p(P+\overline{P}) < (\overline{p}+p)\overline{P} = \overline{P}$، یعنی این‌که $p < \min(P, \overline{P})$. صرف‌نظر از این‌که ۰ یا ۱ دریافت شود با احتمال بیشتر ۱ ارسال شده است.

(ب) $pP < \overline{p}\,\overline{P}$ و $p\overline{P} < \overline{p}P$، یا معادلاً $\overline{P} < p < P$. هر سمبلی که دریافت شود با احتمال بیشتر همان سمبل ارسال شده است.

(ج) $pP < \overline{p}\,\overline{P}$ و $p\overline{P} > \overline{p}P$، یا معادلاً $P < p < \overline{P}$. اگر ۰ یا ۱ دریافت شود آنگاه به‌ترتیب با احتمال بیشتر ۱ یا ۰ ارسال شده است.

۳.۴ با استفاده از $R_{ij} = q_j Q_{ij}$ و $\sum_i R_{ij} = q_j$ داریم

$$\begin{aligned} H(\mathcal{A},\mathcal{B}) &= \sum_i \sum_j R_{ij} \log \frac{1}{R_{ij}} = \sum_i \sum_j R_{ij} \log \frac{1}{q_j} + \sum_i \sum_j R_{ij} \log \frac{1}{Q_{ij}} \\ &= \sum_j q_j \log \frac{1}{q_j} + \sum_i \sum_j R_{ij} \log \frac{1}{Q_{ij}} = H(\mathcal{B}) + H(\mathcal{A}|\mathcal{B}). \end{aligned}$$

پس $H(\mathcal{A}|\mathcal{B}) = H(\mathcal{A},\mathcal{B}) - H(\mathcal{B})$ اطلاعاتی است که گیرنده (که $\mathcal{B}$ را می‌داند) با کشف $\mathcal{A}$ به‌دست می‌آورد. معادلاً، این میزان ابهام او درباره $\mathcal{A}$ به فرض دانستن $\mathcal{B}$ است.

۴.۴ بنابر لم ۲۱.۳، اگر $\mathcal{S}$ و $\mathcal{T}$ دو منبع مستقل باشند آنگاه $H(\mathcal{S}\times\mathcal{T}) = H(\mathcal{S}) + H(\mathcal{T})$. اگر کانال‌های Γ و Γ' به‌ترتیب دارای ورودی‌های $\mathcal{A}$ و $\mathcal{A}'$ و خروجی‌های $\mathcal{B}$ و $\mathcal{B}'$ باشند، از

این بی‌درنگ نتیجه می‌شود $H(\mathcal{A}\times\mathcal{A}')=H(\mathcal{A})+H(\mathcal{A}')$، و مشابهاً این رابطه برای $H(\mathcal{B}\times\mathcal{B}')$ و $H(\mathcal{A}\times\mathcal{A}',\mathcal{B}\times\mathcal{B}')$ برقرار است. اگر b_j و b'_k سمبل‌های نوعی خروجی Γ و Γ' با احتمال‌های q_j و q'_k باشند آنگاه با درنظر گرفتن $\sum_j q_j=\sum_k q'_k=1$ داریم:

$$\begin{aligned}H(\mathcal{A}\times\mathcal{A}'\mid\mathcal{B}\times\mathcal{B}') &= \sum_{j,k} q_jq'_kH(\mathcal{A}\times\mathcal{A}'\mid b_jb'_k)\\ &= \sum_{j,k} q_jq'_k\,(H(\mathcal{A}|b_j)+H(\mathcal{A}'|b'_k))\\ &= \sum_j q_jH(\mathcal{A}|b_j)+\sum_k q'_kH(\mathcal{A}'|b'_k)\\ &= H(\mathcal{A}|\mathcal{B})+H(\mathcal{A}'|\mathcal{B}').\end{aligned}$$

یک بحث مشابه نشان می‌دهد $H(\mathcal{B}\times\mathcal{B}'\mid\mathcal{A}\times\mathcal{A}')=H(\mathcal{B}\mid\mathcal{A})+H(\mathcal{B}'\mid\mathcal{A}')$. نتایج متناظر برای Γ^n با استقرا روی n به‌دست می‌آیند.

۵.۴ فرض کنید نتیجه غلط باشد، پس برای عددی چون $c=\lambda a+\overline{\lambda}b$ که $a<b$ و $0<\lambda<1$، داریم $f(c)\le\lambda f(a)+\overline{\lambda}f(b)$ (پس $a<c<b$). از به‌کار بردن قضیه مقدار میانی روی f در بازه‌های $[a,c]$ و $[c,b]$ نتیجه می‌شود برای اعدادی چون c_1 و c_2، که $a<c_1<c<c_2<b$، داریم $f'(c_1)=(f(c)-f(a))/(c-a)$ و $f'(c_2)=(f(b)-f(c))/(b-c)$. از جایگزینی برای c و به‌کار بردن نامساوی مربوط به $f(c)$ نتیجه می‌شود

$$\begin{aligned}f'(c_1)\le\frac{\lambda f(a)+\overline{\lambda}f(b)-f(a)}{\lambda a+\overline{\lambda}b-a} &= \frac{f(b)-f(a)}{b-a}\\ &= \frac{f(b)-\lambda f(a)-\overline{\lambda}f(b)}{b-\lambda a-\overline{\lambda}b}\le f'(c_2),\end{aligned}$$

و بنابراین از قضیه مقدار میانی (اعمال شده بر f' روی $[c_1,c_2]$) نتیجه می‌شود برای عددی چون c_3 که $c_1<c_3<c_2$ داریم

$$f''(c_3)=(f'(c_2)-f'(c_1))/(c_2-c_1)\ge 0.$$

این متناقض با منفی بودن f'' روی $(0,1)$ است.

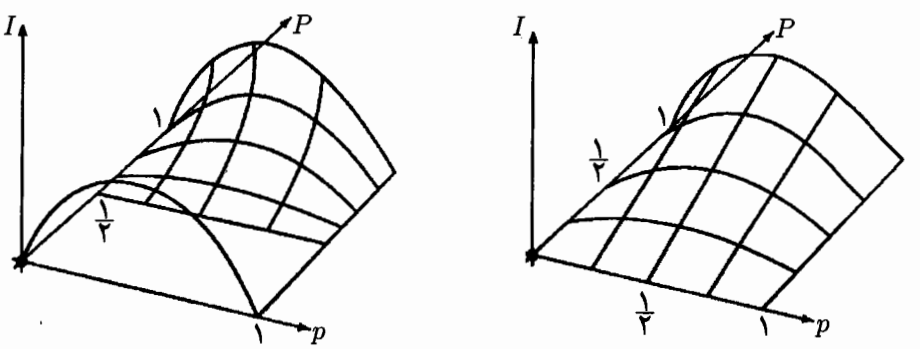
۶.۴ فرض کنید Γ یک کانال دوتایی با ماتریس $\begin{pmatrix}1&0\\1&0\end{pmatrix}$ باشد، پس هر سمبل $a\in\{0,1\}$ به‌صورت $b=0$ دریافت می‌شود. اگر احتمال‌های ورودی برابر p و $\bar{p}$ باشند ($p=Pr(a=0)$) آنگاه $H(\mathcal{A})=H(p)$، درحالی‌که $H(\mathcal{B})=H(1)=0$، چون احتمال‌های خروجی برابر ۱ و ۰ هستند. پس اگر $0<p<1$ آنگاه $H(\mathcal{A})>H(\mathcal{B})$.

۷.۴ سمبل‌های ورودی ۰ و ۱ دارای احتمال‌های p و $\bar{p}$ هستند، پس $H(\mathcal{A}) = H(p)$. سمبل‌های خروجی ۰، ۱، و ؟ دارای احتمال‌های pP، $\bar{p}P$ و $\overline{P}$ هستند، بنابراین

$$\begin{aligned}
H(\mathcal{B}) &= -pP\log pP - \bar{p}P\log \bar{p}P - \overline{P}\log\overline{P} \\
&= -P(p\log p + \bar{p}\log\bar{p} + \log P) - \overline{P}\log\overline{P} = PH(p) + H(P), \\
H(\mathcal{B}|\mathcal{A}) &= -pP\log P - p\overline{P}\log\overline{P} - \bar{p}P\log P - \bar{p}\overline{P}\log\overline{P} \\
&= -P\log P - \overline{P}\log\overline{P} = H(P), \\
H(\mathcal{A},\mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}|\mathcal{A}) = H(p) + H(P) \quad \text{((بنابر ۶.۴))}, \\
H(\mathcal{A}|\mathcal{B}) &= H(\mathcal{A},\mathcal{B}) - H(\mathcal{B}) = H(p) - PH(p) = \overline{P}H(p) \quad \text{((بنابر ۷.۴))}.
\end{aligned}$$

سپس از $P, H(p) \geq 0$ نتیجه می‌شود $H(\mathcal{B}) \geq H(\mathcal{B}|\mathcal{A})$، و $\overline{P} \geq 1$ ایجاب می‌کند $H(\mathcal{A}) \geq H(\mathcal{A}|\mathcal{B})$.

۸.۴



۹.۴ بنابر تمرین ۷.۴، برای کانال BEC داریم $I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}) = PH(p)$. با ثابت بودن P و متغیر بودن p، $I(\mathcal{A},\mathcal{B})$ ماکزیمم خود را به‌ازای $p = \frac{1}{2}$ $(H(p) = 1)$ اختیار می‌کند و بنابراین $C = I_{\max} = P$.

۱۰.۴ اگر Γ و Γ' به‌ترتیب ورودی‌های $\mathcal{A}$ و $\mathcal{A}'$ و خروجی‌های $\mathcal{B}$ و $\mathcal{B}'$ داشته باشند، آنگاه از تمرین ۴.۴ نتیجه می‌شود

$$H(\mathcal{B}\times\mathcal{B}') = H(\mathcal{B}) + H(\mathcal{B}'), \quad H(\mathcal{B}\times\mathcal{B}' \mid \mathcal{A}\times\mathcal{A}') = H(\mathcal{B}\mid\mathcal{A}) + H(\mathcal{B}'\mid\mathcal{A}');$$

پس $I(\mathcal{A}\times\mathcal{A}', \mathcal{B}\times\mathcal{B}') = I(\mathcal{A},\mathcal{B}) + I(\mathcal{A}',\mathcal{B}')$، و با ماکزیمم گرفتن روی تمامی $\mathcal{A}$ و $\mathcal{A}'$ ملاحظه می‌شود که $\Gamma\times\Gamma'$ دارای ظرفیت $C + C'$ است. با استقرا روی n دیده می‌شود که Γ^n دارای ظرفیت nC است.

۱۱.۴ اگر $\mathbf{p} = (p_i) \in \mathcal{P}$ آنگاه برای هر $1 \le i \le r$ داریم $0 \le p_i \le 1$، پس $|\mathbf{p}|^2 = \sum_i p_i^2 \le r$؛ از این‌رو $\mathcal{P}$ محدود است. برای نشان دادن این‌که $\mathcal{P}$ بسته است، فرض کنید $\mathbf{y} = (y_i) \in \mathbf{R}^r \backslash \mathcal{P}$، پس یا یکی از y_i ها منفی بوده و یا $\sum y_i \ne 1$. در حالت اول، تمامی بردارهای $\mathbf{x} \in \mathbf{R}^r$ با خاصیت $|\mathbf{x} - \mathbf{y}| < |y_i|$ در شرط $x_i < 0$ صدق می‌کنند زیرا $|x_i - y_i| \le |\mathbf{x} - \mathbf{y}|$، پس $\mathbf{x} \notin \mathcal{P}$. در حالت دوم، $\mathbf{y}$ دارای فاصله $d = |\sum y_i - 1|/\sqrt{r}$ از فوق صفحه $\sum p_i = 1$ است، و بنابراین تمامی $\mathbf{x}$ های با خاصیت $|\mathbf{x} - \mathbf{y}| < d$ در خارج $\mathcal{P}$ قرار دارند.

۱۲.۴ Γ دارای ماتریس کانال $M = \begin{pmatrix} P & Q \\ Q & P \end{pmatrix}$ است که $Q = \overline{P} = 1 - P$، پس بنابر تمرین ۱.۴ کانال Γ^n دارای ماتریس M^n است. با استقرا روی n ملاحظه می‌شود که M^n به‌فرم $\begin{pmatrix} P_n & Q_n \\ Q_n & P_n \end{pmatrix}$ است که $0 \le P_n \le 1$ و $Q_n = \overline{P}_n$، از این‌رو Γ^n یک BSC است. حال M دارای مقادیر ویژه $\lambda = 1, 2P - 1$ است، پس M^n دارای مقادیر ویژه $1^n, (2P-1)^n$ است؛ بنابراین $2P_n = tr(M^n) = 1 + (2P-1)^n$، و از این‌رو $P_n = (1 + (2P-1)^n)/2$ و $Q_n = (1 - (2P-1)^n)/2$ (این روابط را می‌توانید با استقرا نیز ثابت کنید). پس Γ^n دارای ظرفیت C_n است:

$$C_n = 1 - H(P_n) = 1 - H((1 + (2P-1)^n)/2).$$

با افزایش n و با شرط $0 < P < 1$ ملاحظه می‌شود $(2P-1)^n \longrightarrow 0$، از این‌رو C_n به $1 - H(\frac{1}{2}) = 0$ میل می‌کند. اگر P برابر 0 و 1 باشد آنگاه هر P_n نیز برابر 0 یا 1 بوده و بنابراین $C_n = 1$.

۱۳.۴ $C = I_{\max}$، پس $C = 0$ اگر و فقط اگر برای هر $\mathcal{A}$ داشته باشیم $I(\mathcal{A}, \mathcal{B}) = 0$، یعنی این‌که (بنابر قضیه ۱۱.۴) $\mathcal{A}$ و $\mathcal{B}$ به‌ازای هر $\mathcal{A}$ مستقل هستند. این بدین معنی است که برای هر i و j داریم $P_{ij} = Pr(b_j|a_i) = Pr(b_j)$، یعنی این‌که تمامی سطرهای M برابر هستند. تفسیر موضوع این است که توزیع احتمال ورودی تأثیری روی توزیع خروجی ندارد، بنابراین گیرنده اطلاعاتی راجع به ورودی به‌دست نمی‌آورد.

۱۴.۴ در صورت لزوم می‌توان آنتروپی H را در یک ثابت ضرب کرد، از این‌رو می‌توانیم فرض کنیم $r = e$، پس $H(\mathbf{x}) = -\sum_i x_i \ln x_i$ و در نتیجه برای هر $x_i > 0$ داریم $\partial H/\partial x_i = -1 - \ln x_i$. اگر در $\mathcal{P}$ داشته باشیم $\mathbf{p} \ne \mathbf{q}$ آنگاه تابع $f(\lambda) = H(\lambda \mathbf{p} + \overline{\lambda}\mathbf{q})$ روی $[0, 1]$ پیوسته است و برای هر $\lambda \in (0, 1)$ داریم

$$\begin{aligned} f'(\lambda) &= -\sum_i (1 + \ln(\lambda p_i + \overline{\lambda} q_i))(p_i - q_i), \\ f''(\lambda) &= -\sum_i (p_i - q_i)^2/(\lambda p_i + \overline{\lambda} q_i), \end{aligned}$$

که عمل جمع روی i های با خاصیت $\lambda p_i + \overline{\lambda} q_i > 0$ انجام شده است. پس تابع f'' روی $(0,1)$ منفی بوده و بنابراین بنابر لم ۶.۴ تابع f روی $[0,1]$ اکیداً محدب است. در نتیجه برای هر $\lambda \in [0,1]$ داریم $H(\lambda \mathbf{p} + \overline{\lambda}\mathbf{q}) \geq \lambda H(\mathbf{p}) + \overline{\lambda} H(\mathbf{q})$، و تساوی برقرار است اگر و فقط اگر $\lambda \in \{0,1\}$.

۱۵.۴ $I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A})$ و $H(\mathcal{B} \mid \mathcal{A}) = \sum_i p_i \left(\sum_j P_{ij} \log P_{ij} \right)$. از شرط روی سطرها نتیجه می‌شود که $\sum_j P_{ij} \log P_{ij}$ برابر یک ثابت c مستقل از i می‌باشد، پس $I(\mathcal{A},\mathcal{B}) = H(\mathcal{B}) + c$ زیرا $\sum_i p_i = 1$. حال c مستقل از $\mathcal{A}$ است، پس ماکزیمم کردن $I(\mathcal{A},\mathcal{B})$ معادل با ماکزیمم کردن $H(\mathcal{B})$ است. بنابر قضیه ۱۰.۳، ماکزیمم $H(\mathcal{B})$ برابر $\log s$ است، از این‌رو $C = I_{\max} = \log s + c$ که وقتی تمامی q_i ها برابر باشند حاصل می‌شود؛ چون $q_j = \sum_i p_i P_{ij}$، از شرط روی ستون‌ها نتیجه می‌شود که این رخ می‌دهد اگر و فقط اگر تمامی p_i ها برابر باشند. برای کانال متقارن r تایی به ظرفیت C می‌رسیم:

$$C = \log s + c = \log r + P \log P + \overline{P} \log \overline{P} - \overline{P} \log(r-1).$$

برای $r = 2$ این هماهنگ با مقدار $1 - H(P)$ برای کانال BSC می‌باشد.

۱۶.۴

$$\begin{aligned} I(\mathcal{A},\mathcal{B}) &= H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A}) \\ &= -q_1 \log q_1 - q_2 \log q_2 + p_1(P_{11} \log P_{11} + P_{12} \log P_{12}) \\ &+ p_2(P_{21} \log P_{21} + P_{22} \log P_{22}). \end{aligned}$$

دو معادله خطی $P_{i1}c_1 + P_{i2}c_2 = P_{i1} \log P_{i1} + P_{i2} \log P_{i2}$ برای c_1 و c_2 قابل حل است اگر $\det(P_{ij}) \neq 0$، یا معادلاً برای $j = 1,2$ داشته باشیم $P_{1j} \neq P_{2j}$، و اگر این شرط برقرار نباشد هنوز با قرار دادن $c_j = \log P_{1j} = \log P_{2j}$ قابل حل هستند. در این‌صورت

$$I = -q_1 \log q_1 - q_2 \log q_2 + p_1(P_{11}c_1 + P_{12}c_2) + p_2(P_{21}c_1 + P_{22}c_2)$$

و چون برای $j = 1,2$ داریم $p_1 P_{1j} + p_2 P_{2j} = q_j$، پس I را به‌عنوان تابعی از q_1 و q_2 داریم

$$I = -q_1 \log q_1 - q_2 \log q_2 + q_1 c_1 + q_2 c_2.$$

برای ماکزیمم نمودن I تحت شرط $q_1 + q_2 = 1$، تعریف کنید $\Phi = I + \lambda(q_1 + q_2 - 1)$ و معادله $\partial\Phi/\partial q_1 = \partial\Phi/\partial q_2 = q_1 + q_2 - 1 = 0$ را حل کنید. از دو معادله اول نتیجه می‌شود $c_j + \lambda = 1 + \log q_j$، و بنابراین $c_1 - \log q_1 = c_2 - \log q_2$. در این‌صورت با استفاده از $q_1 + q_2 = 1$ برای $j = 1, 2$ داریم

$$C = I_{\max} = q_1(c_1 - \log q_1) + q_2(c_2 - \log q_2) = c_j - \log q_j.$$

پس $2^{c_j} = 2^C q_j$، و از این‌رو $2^C = 2^C(q_1+q_2) = 2^{c_1} + 2^{c_2}$، پس $C = \log(2^{c_1} + 2^{c_2})$. اگر $P_{11} = P_{22}$ آنگاه Γ کانال BSC بوده با $P_{11} = P_{22} = P$ و $P_{12} = P_{21} = \overline{P}$؛ از معادلات خطی $Pc_1 + \overline{P}c_2 = -H(P) = \overline{P}c_1 + Pc_2$ نتیجه می‌شود $c_1 = c_2 = -H(P)$ و $C = \log(2^{c_1} + 2^{c_2}) = 1 - H(P)$.

۱۷.۴ فرض کنید کانال‌های Γ_1، Γ_2، و $\Gamma = \Gamma_1 + \Gamma_2$ به‌ترتیب دارای r، r' و $r + r'$ سمبل ورودی و s، s' و $s + s'$ سمبل خروجی باشند. فرض کنید $(p_1, \ldots, p_{r+r'})$ یک توزیع ورودی برای Γ باشد که در آن سمبل‌های Γ_1 قبل از سمبل‌های Γ_2 قرار گرفته‌اند. اگر $u = p_1 + \cdots + p_r$ و $v = p_{r+1} + \cdots + p_{r+r'}$، پس $u + v = 1$، آنگاه $(p_1/u, \ldots, p_r/u)$ و $(p_{r+1}/v, \ldots, p_{r+r'}/v)$ به‌ترتیب توزیع‌های ورودی برای Γ_1 و Γ_2 می‌باشند. اگر از (p_i) توزیع خروجی (q_j) برای Γ به‌دست آید، آنگاه به جهت خطی بودن، از $(p_1/u, \ldots, p_r/u)$ توزیع خروجی $(q_1/u, \ldots, q_s/u)$ برای Γ_1 نتیجه می‌شود؛ بالاخص $q_1 + \cdots + q_s = u$. خروجی $\mathcal{B}_1$ از Γ_1 دارای آنتروپی

$$H(\mathcal{B}_1) = -\sum_{i=1}^{r}(q_i/u)\log(q_i/u) = \log u - (1/u)\sum_{i=1}^{r} q_i \log q_i$$

است، پس $\sum_{i=1}^{r} q_i \log q_i = u \log u - uH(\mathcal{B}_1)$. از نتیجه مشابه‌ای برای Γ_2 داریم

$$H(\mathcal{B}) = -u \log u - v \log v + uH(\mathcal{B}_1) + vH(\mathcal{B}_2).$$

مشابهاً $H(\mathcal{B}|\mathcal{A}) = uH(\mathcal{B}_1|\mathcal{A}_1) + vH(\mathcal{B}_2|\mathcal{A}_2)$، پس

$$I(\mathcal{A}, \mathcal{B}) = -u \log u - v \log v + uI(\mathcal{A}_1, \mathcal{B}_1) + vI(\mathcal{A}_2, \mathcal{B}_2).$$

برای ماکزیمم کردن $I(\mathcal{A}, \mathcal{B})$ قرار می‌دهیم $I(\mathcal{A}_i, \mathcal{B}_i) = C_i$ (بیشترین مقدار آن) و سپس u و v را برای ماکزیمم کردن $I = -u \log u - v \log v + uC_1 + vC_2$ با شرط $u + v = 1$ تغییر می‌دهیم. این اساساً مسئله‌ای است که در تمرین ۱۶.۴ داشتیم، پس از روش به‌کار رفته در آن‌جا نتیجه می‌شود $C = \log(2^{C_1} + 2^{C_2})$. وقتی $\Gamma_1 = \Gamma_2$ نتیجه می‌گیریم $C = C_1 + 1$.

۱۸.۴ $Pr(a|c) = \sum_b Pr(a|b)Pr(b|c)$، پس از ضرب کردن در $Pr(c)$ و به‌کار بردن $Pr(c)Pr(b|c) = Pr(b, c)$ نتیجه می‌شود $Pr(c)Pr(a|c) = \sum_b Pr(a|b)Pr(b, c)$، و

از این‌رو

$$\sum_b \sum_c \left(Pr(b,c) \sum_a Pr(a|b) \log Pr(a|c) \right) = \sum_c \left(Pr(c) \sum_a Pr(a|c) \log Pr(a|c) \right) = -H(\mathcal{A}|\mathcal{C}).$$

نیز از $\sum_c Pr(b,c) = Pr(b)$ نتیجه می‌شود

$$\sum_b \sum_c \left(Pr(b,c) \sum_a Pr(a|b) \log Pr(a|b) \right) = \sum_b \left(Pr(b) \sum_a Pr(a|b) \log Pr(a|b) \right) = -H(\mathcal{A}|\mathcal{B}),$$

بنابراین

$$\sum_b \sum_c \left(Pr(b,c) \sum_a Pr(a|b)(\log Pr(a|b) - \log Pr(a|c)) \right) = H(\mathcal{A}|\mathcal{C}) - H(\mathcal{A}|\mathcal{B}).$$

نتیجه ۹.۳ نشان می‌دهد که رابطه $\sum_a Pr(a|b)\,(\log Pr(a|b) - \log Pr(a|c)) \geq 0$ برای هر b و c برقرار است، پس $H(\mathcal{A}|\mathcal{C}) \geq H(\mathcal{A}|\mathcal{B})$ و از این‌رو

$$I(\mathcal{A},\mathcal{C}) = H(\mathcal{A}) - H(\mathcal{A} \mid \mathcal{C}) \leq H(\mathcal{A}) - H(\mathcal{A} \mid \mathcal{B}) = I(\mathcal{A},\mathcal{B}).$$

این نامساوی‌ها نشان می‌دهد که ارسال بیشتر (از $\mathcal{B}$ به $\mathcal{C}$) هرگز ابهام راجع به $\mathcal{A}$ را کاهش نداده و اطلاعات متقابل راجع به $\mathcal{A}$ را نیز هرگز افزایش نمی‌دهد. داریم $C = \max I(\mathcal{A},\mathcal{C}) \leq \max I(\mathcal{A},\mathcal{B}) = C_1$، و مشابهاً از $I(\mathcal{A},\mathcal{C}) \leq I(\mathcal{B},\mathcal{C})$ نتیجه می‌شود $C \leq C_2$، پس $C \leq \min(C_1, C_2)$. اگر $\Gamma_1 = \Gamma_2$ یک کانال BSC با ظرفیت $C_1 = C_2 = 1 - H(P)$ باشد، آنگاه تمرین ۱۲.۴ نشان می‌دهد که Γ یک کانال BSC با احتمال $P' = (1 + (2P-1)^2)/2$ و ظرفیت $C = 1 - H(P')$ است. اگر $P \in \{0, 1\}$ آنگاه $P' = 1$ و $C = C_1 = C_2 = 1$؛ اگر $P = \frac{1}{2}$ آنگاه $P' = \frac{1}{2}$ و $C = C_1 = C_2 = 0$. در غیراین‌صورت، $|P - \frac{1}{2}| < |P' - \frac{1}{2}|$ و از این‌رو $C < C_1 = C_2$.

۱.۵ یک قاعده تصمیم صرفاً یک تابع $\Delta : B \longrightarrow A$ است، پس تعداد $|A|^{|B|} = r^s$ قاعده تصمیم وجود دارد.

۲.۵ داریم

$$(R_{ij}) = \begin{pmatrix} pP & p\overline{P} \\ \bar{p}\overline{P} & \bar{p}P \end{pmatrix} = \begin{pmatrix} 0.72 & 0.18 \\ 0.02 & 0.08 \end{pmatrix},$$

و بزرگ‌ترین درایه در هر ستون اولین درایه است، پس $\Delta(۰) = \Delta(۱) = ۰$، و بنابراین

$$Pr_E = ۱ - Pr_C = ۱ - (۰.۷۲ + ۰.۱۸) = ۰.۱.$$

۳.۵ برای هر قاعده تصمیم $\Delta : B \longrightarrow A$، $b_j \longmapsto a_i = a_{j^\bullet}$،

$$\int_{\mathbf{p}\in\mathcal{P}} Pr_C dp = \int_{\mathbf{p}\in\mathcal{P}} \left(\sum_j p_{j^\bullet} P_{j^\bullet j}\right) dp = \sum_j \left(P_{j^\bullet j} \int_{\mathbf{p}\in\mathcal{P}} p_{j^\bullet} dp\right),$$

زیرا در حین تغییر $\mathbf{p}$ هر $P_{j^\bullet j}$ ثابت است. حال $\int_{\mathbf{p}\in\mathcal{P}} p_{j^\bullet} dp$ دارای مقدار یکسان i برای تمامی j و Δ دارد، زیرا $\mathcal{P}$ تحت تمامی جایگشت‌های مختصات p_i متقارن است. در نتیجه Δ مقدار $\int_{\mathbf{p}\in\mathcal{P}} Pr_C dp$ را ماکزیمم می‌کند اگر مقدار $P_{j^\bullet j}$ را برای هر j ماکزیمم کند و این قاعده بیشترین درستنمایی است.

۴.۵ $d(\mathbf{u},\mathbf{v}) = i$ اگر و فقط اگر $\mathbf{v}$ و $\mathbf{u}$ در دقیقاً i موقعیت مختصاتی با هم متفاوت باشند؛ تعداد $\binom{n}{i}$ حالت ممکن برای انتخاب یک چنین موقعیت‌هایی وجود دارد، و برای هر مختصی $\mathbf{v}$ می‌تواند تعداد $r-۱$ مقدار متفاوت داشته باشد، پس تعداد $\binom{n}{i}(r-۱)^i$ امکان برای $\mathbf{v}$ وجود دارد. از قضیه دوجمله‌ای نتیجه می‌شود

$$\sum_{i=۰}^{n} \binom{n}{i}(r-۱)^i = (r-۱+۱)^n = r^n = |A^n|.$$

۵.۵ بزرگ‌ترین زیرمجموعه‌های با این خاصیت چهار عضو دارند. این‌ها مجموعه رئوس $\{۰۰۰, ۱۱۰, ۱۰۱, ۰۱۱\}$ و $\{۱۰۰, ۰۱۰, ۰۰۱, ۱۱۱\}$ از دو چهاروجهی محاط شده در مکعب $\mathbf{Z}_۲^۳$ هستند. در $\mathbf{Z}_۲^n$ بزرگ‌ترین یک چنین مجموعه‌ای دارای $۲^{n-۱}$ عضو است: دو مجموعه این‌چنینی وجود دارد، حاوی کلمات به طول n با تعداد زوجی، یا تعداد فردی، از سمبل 1.

۶.۵ فرض کنید $\mathbf{u},\mathbf{v},\mathbf{w} \in A^n$. اگر $u_i \neq w_i$، آنگاه $u_i \neq v_i$ یا $v_i \neq w_i$، پس

$$\begin{aligned} d(\mathbf{u},\mathbf{w}) &= |\{i|u_i \neq w_i\}| \\ &\leq |\{i|u_i \neq v_i \text{ یا } v_i \neq w_i\}| \\ &\leq |\{i|u_i \neq v_i\}| + |\{i|v_i \neq w_i\}| \\ &= d(\mathbf{u},\mathbf{v}) + d(\mathbf{v},\mathbf{w}). \end{aligned}$$

۷.۵ چون $\lambda + \mu = ۱$، از قضیه دوجمله‌ای نتیجه می‌شود

$$۱ = (\lambda+\mu)^n = \sum_{i=۰}^{n} \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i \leq \lambda n} \binom{n}{i} \lambda^i \mu^{n-i} \geq \sum_{i \leq \lambda n} \binom{n}{i} \lambda^{\lambda n} \mu^{\mu n};$$

آخرین نامساوی به دلیل $\lambda/\mu \leq ۱$ بوده و از $i \leq \lambda n$ نتیجه می‌شود

$$\lambda^i \mu^{n-i} = (\frac{\lambda}{\mu})^i \mu^n \geq (\frac{\lambda}{\mu})^{\lambda n} \mu^n = \lambda^{\lambda n} \mu^{n-\lambda n} = \lambda^{\lambda n} \mu^{\mu n}.$$

از تقسیم کردن بر $\lambda^{\lambda n}\mu^{\mu n}$ نتیجه می‌شود

$$\sum_{i \leq \lambda n} \binom{n}{i} \leq \lambda^{-\lambda n} \mu^{-\mu n} = (\lambda^{-\lambda}\mu^{-\mu})^n,$$

پس

$$\log_۲ \sum_{i \leq \lambda n} \binom{n}{i} \leq n(-\lambda \log_۲ \lambda - \mu \log_۲ \mu) = n H_۲(\lambda)$$

و بنابراین $\sum_{i \leq \lambda n} \binom{n}{i} \leq ۲^{n H_۲(\lambda)}$.

۸.۵ $\mathcal{R}_n = \{\mathbf{0} = ۰۰\ldots۰, \mathbf{1} = ۱۱\ldots۱\}$. کلمه دریافتی $\mathbf{v}$ متشکل از n سمبل مساوی ۰ و ?، و یا ۱ و ? است زیرا $\mathbf{u} \in \{\mathbf{0},\mathbf{1}\}$ ارسال شده است، پس فرض کنید $\Delta(\mathbf{v}) = \mathbf{0},\mathbf{1}$ اگر کلمه دریافتی $\mathbf{v}$ حاوی یک سمبل ۰ یا ۱ باشد، و فرض کنید $\Delta(\mathbf{v})$ برای $\mathbf{v} = ??\ldots?$ تعریف نشده است. در این‌صورت کدگشایی درست است مگر $\mathbf{v} = ??\ldots?$، پس $Pr_E = Pr(\mathbf{v} = ??\ldots?) = \overline{P}^n \longrightarrow ۰$.

۹.۵ ماتریس کانال $\begin{pmatrix} ۱ & ۰ \\ Q & P \end{pmatrix}$ است. چون $Q > ۱$ و $P > ۰$، قاعده بیشترین درستنمایی به‌صورت $\Delta(۰) = ۰$ و $\Delta(۱) = ۱$ است با $Pr_C = p + \overline{p}P$ و $Pr_E = \overline{p}Q$. اگر ۰۰۰ ارسال شود درست دریافت می‌شود. اگر ۱۱۱ ارسال شود با ۰، ۱، ۲ یا ۳ خطا دریافت می‌شود، به‌ترتیب با احتمال‌های $P^۳, P^۲Q, ۳P^۲Q, ۳PQ^۲, Q^۳$. چون $Q^۳ > ۱$ و $P^۳, ۳P^۲Q, ۳Q^۲P > ۰$، از قاعده بیشترین درستنمایی داریم $\Delta(۰۰۰) = ۰۰۰$ و $\Delta(\mathbf{v}) = ۱۱۱$ اگر $\mathbf{v} \neq ۰۰۰$. این متفاوت با قاعده اکثریت و قاعده نزدیک‌ترین همسایه است، زیرا (برای مثال) $\Delta(۱۰۰) = ۱۱۱ \neq ۰۰۰$. از قاعده بیشترین درستنمایی نتیجه می‌شود $Pr_C = p + \overline{p}(۱ - Q^۳)$، $Pr_E = \overline{p}Q^۳$، و نرخ برابر $R = ۱/۳$ است. اگر $\mathcal{R}_n = \{۰۰\ldots۰, ۱۱\ldots۱\}$ به‌کار رود، $Pr_E = \overline{p}Q^n$ و $R = ۱/n$ که هر دو به صفر میل می‌کنند.

۱۰.۵ اگر d معرف $d(\mathbf{u},\mathbf{v})$ باشد، احتمال به‌جلو برابر $Pr(\mathbf{v}|\mathbf{u}) = P^{n-d}Q^d = P^n(Q/P)^d$ است؛ چون $Q/P < 1$ این با افزایش d کم می‌شود، پس قاعده بیشترین درستنمایی (برای $\mathbf{v}$ مفروض، $Pr(\mathbf{v}|\mathbf{u})$ ماکزیمم می‌شود) d را می‌نیمم می‌کند، و از این‌رو با کدگشایی نزدیک‌ترین همسایه هماهنگی دارد. اگر w معرف $d(\mathbf{0},\mathbf{v})$ باشد آنگاه $d(\mathbf{1},\mathbf{v}) = n - w$، پس بسته به $w < n - w$ یا $w > n - w$ مقدار $\Delta(\mathbf{v})$ برابر $\mathbf{0}$ یا $\mathbf{1}$ خواهد بود؛ حال $\mathbf{v}$ دارای w سمبل $v_i = 1$ و $n - w$ سمبل $v_i = 0$ است، پس Δ با کدگشایی اکثریت برابر است. با به‌کار بردن این قاعده Δ، و قرار دادن $n = 2t + 1$، داریم

$$\begin{aligned}
Pr_E &= Pr\ (\text{بیشتر از } t \text{ خطا}) \\
&= \binom{2t+1}{t+1}P^tQ^{t+1} + \binom{2t+1}{t+2}P^{t-1}Q^{t+2} + \cdots + \binom{2t+1}{2t+1}P^0Q^{2t+1} \\
&\le (t+1)\binom{2t+1}{t+1}P^tQ^{t+1} \\
&= \frac{(2t+1)!}{(t!)^2}P^tQ^{t+1}\ (a_t \text{ مساوی فرضا})
\end{aligned}$$

زیرا در $t+1$ جمله موجود بزرگ‌ترین آنها اولین است $(Q/P < 1)$ و ضرایب دوجمله‌ای کاهش می‌یابند. چون اگر $P \le 1$ آنگاه $P < \frac{1}{2}$ $PQ = P - P^2 < \frac{1}{4}$ با افزایش t داریم

$$\frac{a_{t+1}}{a_t} = \frac{(2t+3)(2t+2)}{(t+1)^2}PQ \longrightarrow 4PQ < 1,$$

پس اگر $t \longrightarrow \infty$ آنگاه $a_t \longrightarrow 0$، و بنابراین با افزایش n مقدار Pr_E به صفر میل می‌کند. نرخ $R = \frac{1}{n}$ به 0 میل می‌کند درحالی‌که قضیه شانون نیاز به $R \approx C > 0$ دارد، پس این قضیه را ثابت نمی‌کند.

۱۱.۵ بسته به این‌که Γ نتیجه را درست یا نادرست انتقال دهد، با هر پرتاب موجودی حاضر در 2λ یا 2μ ضرب می‌شود، پس بعد از m ارسال درست و $n - m$ ارسال نادرست سرمایه اولیه در $(2\lambda)^m(2\mu)^{n-m} = 2^n\lambda^m\mu^{n-m}$ ضرب می‌شود. از این‌رو $c_n = 2^n\lambda^m\mu^{n-m}c_0$، و بنابراین $\frac{1}{n}\log(c_n/c_0) = 1 + \frac{m}{n}\log\lambda + \frac{n-m}{n}\log\mu$. بنابر قانون اعداد بزرگ، می‌توان انتظار داشت $m/n \approx P$ و $(n-m)/n \approx Q$، با احتمال متمایل به ۱ وقتی n به ∞ میل می‌کند؛ پس $G \approx 1 + P\log\lambda + Q\log\mu$. ماکزیمم کردن G معادل با انتخاب λ، μ برای می‌نیمم کردن $-P\log\lambda - Q\log\mu$ است، و بنابر نتیجه ۹.۳ این با انتخاب $\lambda = P$ (به‌قسمی‌که $\mu = Q$) حاصل می‌شود و $G \approx 1 + P\log P + Q\log Q = 1 - H(P) = C$. اگر $P < \frac{1}{2}$ آنگاه استفاده از یک کد تکراری باعث کاهش احتمال خطای Γ و بنابراین افزایش C و G می‌شود.

۱۲.۵ اگر b_j دریافت شود، قمارباز برای هر a_i روی λ_{ij} درصد سرمایه خود شرط می‌بندد، که $\sum_i \lambda_{ij} = ۱$. اگر ورودی برابر a_i باشد این سرمایه را در λ_{ij}/p_i ضرب می‌کند، پس بعد از n شرط داریم $c_n = \prod_i \prod_j (\lambda_{ij}/p_i)^{m_{ij}} c_۰$ که m_{ij} تعداد دفعاتی است که a_i ارسال و b_j دریافت می‌شود. بنابراین

$$G = \lim_{n\to\infty} \frac{۱}{n} \log(c_n/c_۰) = \sum_i \sum_j \lim_{n\to\infty} (m_{ij}/n) \log(\lambda_{ij}/p_i).$$

بنابر قانون اعداد بزرگ با میل کردن n به ∞ با احتمال متمایل به ۱ داریم $m_{ij}/n \approx R_{ij}$، پس

$$\begin{aligned} G \approx \sum_i \sum_j R_{ij} \log(\lambda_{ij}/p_i) &= \sum_i \sum_j R_{ij} \log \lambda_{ij} - \sum_i p_i \log p_i \\ &= \sum_j \left(\sum_i R_{ij} \log \lambda_{ij} \right) + H(\mathcal{A}). \end{aligned}$$

با داشتن $\mathcal{A}$ و Γ قمارباز می‌تواند با ماکزیمم کردن $\sum_i R_{ij} \log \lambda_{ij}$ ها (برای هر j) G را ماکزیمم کند. چون برای هر j داریم $\sum_i R_{ij} = q_j$، از نتیجه ۹.۳ ملاحظه می‌شود که این با انتخاب $\lambda_{ij} = R_{ij}/q_j = Q_{ij}$ به‌دست می‌آید، پس

$$G \approx \sum_i \sum_j R_{ij} \log Q_{ij} + H(\mathcal{A}) = -H(\mathcal{A}|\mathcal{B}) + H(\mathcal{A}) = I(\mathcal{A}, \mathcal{B}).$$

با تغییر p_i بیشترین مقدار ممکن که این می‌تواند داشته باشد ظرفیت C از Γ است. اگر یک شرط موفق $۱/p'_i$ گرو را به‌دست آورد، λ_{ij}/p_i را با λ_{ij}/p'_i جایگزین می‌کنیم، پس مجدداً $\lambda_{ij} = Q_{ij}$، که بنابر نتیجه ۹.۳ منجر به یک نرخ رشد نمایی $G' \approx -H(\mathcal{A}|\mathcal{B}) - \sum_i p_i \log p'_i \geq I(\mathcal{A}, \mathcal{B})$ می‌شود؛ بنابراین قمارباز معمولاً داراتر می‌شود.

۱.۶ $\mathcal{C} \cap \mathcal{C}'$ و $\mathcal{C} + \mathcal{C}'$ غیرخالی بوده و تحت ترکیبات خطی بسته هستند، از این‌رو خطی می‌باشند. اگر $\mathcal{C} \subset \mathcal{C}'$ یا $\mathcal{C}' \subset \mathcal{C}$ آنگاه $\mathcal{C} \cup \mathcal{C}'$ برابر $\mathcal{C}'$ یا $\mathcal{C}$ است و از این‌رو خطی می‌باشد؛ اگر $\mathcal{C} \not\subseteq \mathcal{C}'$ و $\mathcal{C}' \not\subseteq \mathcal{C}$ آنگاه $\mathcal{C} \cup \mathcal{C}'$ خطی نیست، برای این‌که اگر $c \in \mathcal{C} \backslash \mathcal{C}'$ و $c' \in \mathcal{C}' \backslash \mathcal{C}$، آنگاه $c, c' \in \mathcal{C} \cup \mathcal{C}'$ ولی $c + c' \notin \mathcal{C} \cup \mathcal{C}'$.

۲.۶ اگر $\mathbf{a} = ۱۱۰۱$ آنگاه $\mathbf{u} = ۱۰۱۰۱۰۱$. اگر $\mathbf{v} = ۱۰۱۰۱۱۱$ دریافت شود آنگاه $\mathbf{s} = ۱۱۰$، که معرف ۶ بوده و اشاره به در خطا بودن سمبل ۶ ام دارد، پس $\mathbf{u} = ۱۰۱۰۱۰۱ = \Delta(\mathbf{v})$. اگر $\mathbf{v}' = ۱۰۱۱۱۱۱$ دریافت شود آنگاه $\mathbf{s}' = ۰۱۰$، بنابراین $\mathbf{u} \neq ۱۱۱۱۱۱۱ = \Delta(\mathbf{v}')$.

۳.٦ با درنظر گرفتن تمامی ترکیبات خطی بردارهای مبنا $\mathbf{u}_i$ در مثال ۵.٦، خواهیم داشت

$$\mathcal{H}_۷ = \{۰۰۰۰۰۰۰, ۱۱۱۰۰۰۰, ۱۰۰۱۱۰۰, ۰۱۰۱۰۱۰, ۱۱۰۱۰۰۱, ۰۱۱۱۱۰۰,$$
$$۱۰۱۱۰۱۰, ۰۰۱۱۰۰۱, ۱۱۰۰۱۱۰, ۰۱۰۰۱۰۱, ۱۰۰۰۰۱۱, ۰۰۱۰۱۱۰,$$
$$۱۰۱۰۱۰۱, ۰۱۱۰۰۱۱, ۰۰۰۱۱۱۱, ۱۱۱۱۱۱۱\}.$$

ملاحظه می‌شود که کمترین وزن در میان کلمات ناصفر برابر ۳ است، پس $d = ۳$.

۴.٦ اعضای $\overline{\mathcal{C}}$ عبارت هستند از $\overline{\mathbf{u}} = u_۱ \dots u_{n+۱}$، که $\mathbf{u} = u_۱ \dots u_n \in \mathcal{C}$ و $u_{n+۱} = (u_۱ + \cdots + u_n) \bmod ۲$. پس بسته به این‌که $wt(\mathbf{u})$ زوج یا فرد باشد $wt(\overline{\mathbf{u}})$ برابر $wt(\mathbf{u})$ یا $wt(\mathbf{u}) + ۱$ خواهد بود، پس بنابر لم ۸.٦ کد $\overline{\mathcal{C}}$ به‌ترتیب دارای می‌نیمم فاصله d یا $d + ۱$ خواهد بود. با انتخاب $\mathcal{C} = \mathcal{H}_۷$، با $d = ۳$، تمرین ۳.٦ نشان می‌دهد $\overline{\mathcal{H}}_۷$ دارای کدکلمات زیر است

$$۰۰۰۰۰۰۰۰, ۱۱۱۰۰۰۰۱, ۱۰۰۱۱۰۰۱, ۰۱۰۱۰۱۰۱, ۱۱۰۱۰۰۱۰, ۰۱۱۱۱۰۰۰,$$
$$۱۰۱۱۰۱۰۰, ۰۰۱۱۰۰۱۱, ۱۱۰۰۱۱۰۰, ۰۱۰۰۱۰۱۱, ۱۰۰۰۰۱۱۱, ۰۰۱۰۱۱۰۱,$$
$$۱۰۱۰۱۰۱۰, ۰۱۱۰۰۱۱۰, ۰۰۰۱۱۱۱۰, ۱۱۱۱۱۱۱۱,$$

پس می‌نیمم فاصله آن ۴ است.

۵.٦ هر دو خاصیت معادل با این شرط هستند که برای عددی چون t، هر کلمه به فاصله حداکثر t از یک کدکلمه منحصربه‌فرد است.

٦.٦ $\mathcal{H}_۷$ یک کد کامل با $t = ۱$ است، پس کدگشایی درست است اگر و فقط اگر حداکثر یک خطا رخ داده باشد. احتمال این امر برابر $P^۷ + ۷P^٦Q = -٦P^۷ + ۷P^٦$ است، پس $Pr_E = ۱ + ٦P^۷ - ۷P^٦$. برای مقادیر کم Q، از قضیه دوجمله‌ای نتیجه می‌شود $P^i = (۱ - Q)^i \approx ۱ - iQ + \binom{i}{۲}Q^۲$، پس $Pr_E \approx ۲۱Q^۲$.

۷.٦ بنابر تمرین ۴.٦، کد $\overline{\mathcal{H}}_۷$ دارای $d = ۴$ بوده و بنابراین $t = ۱$ و

$$\sum_{i=۰}^{t} \binom{n}{i}(q-۱)^i = ۱ + \binom{۸}{۱} = ۹.$$

بنابراین $۲^۴ = ۱٦$ کره $S_t(\mathbf{u})$ تنها $۹ \times ۱٦ = ۱۴۴$ بردار از $۲^۸ = ۲۵٦$ بردار $\mathbf{v} \in \mathcal{V} = F_۲^۸$ را می‌پوشانند، پس $\overline{\mathcal{H}}_۷$ کامل نیست.

۸.٦ تقریب استرلینگ $m! \sim (m/e)^m\sqrt{۲\pi m}$ را روی سه فاکتوریل موجود در $\binom{n}{t} = n!/t!(n-t)!$ به‌کار برده و سپس لگاریتم بگیرید.

۹.۶ اگر $d = ۳$ آنگاه بنابر قضیه ۱۵.۶ داریم $t = \lfloor \frac{d-۱}{۲} \rfloor = ۱$، پس با قرار دادن $q = ۳$ در قضیه ۱۵.۶ نتیجه می‌شود $A_۳(n, ۳) \leq \lfloor ۳^n/(۲n+۱) \rfloor$. اگر $n = ۳, ۴, ۵, ۶, ۷, \ldots$ آنگاه $A_۳(n, ۳) \leq ۳, ۹, ۲۲, ۵۶, ۱۴۵, \ldots$. اگر $d = ۴$ آنگاه $t = ۱$، پس همچون مثال ۱۶.۶ از قضیه ۱۵.۶ نتیجه می‌شود $A_۲(n, ۴) \leq \lfloor ۲^n/(n+۱) \rfloor$. اگر $d = ۵$ آنگاه $t = ۲$، و بنابراین $A_۲(n, ۵) \leq \lfloor ۲^n/(۱ + n + \binom{n}{۲}) \rfloor = \lfloor ۲^{n+۱}/(n^۲ + n + ۲) \rfloor$.

۱۰.۶ از مثال ۲۲.۶ نتیجه می‌شود $A_۲(۴, ۳)$ برابر ۲ یا ۳ است. اگر $\mathcal{C} = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ یک کد دوتایی با $n = ۴$ و $d = ۳$ باشد، آنگاه هرکدام از $\mathbf{w}$ و $\mathbf{v}$ حداقل در ۳ موقعیت از چهار مختص خود با $\mathbf{u}$ تفاوت خواهند داشت؛ حداقل ۲ تا از این موقعیت‌های مختصاتی i و j باید یکسان باشند، بنابراین $v_i \neq u_i \neq w_i$ و $v_j \neq u_j \neq w_j$؛ از دوتایی بودن کد نتیجه می‌شود $v_i = w_i$ و $v_j = w_j$، پس $d(\mathbf{v}, \mathbf{w}) \leq ۲$ که در تناقض با $d = ۳$ است. از این‌رو $A_۲(۴, ۳) < ۳$ و بنابراین $A_۲(۴, ۳) = ۲$. کد $\{۰۰۰۰, ۱۱۱۰\}$ این کران را اختیار می‌کند.

۱۱.۶ از قضیه ۱۵.۶ با $d = ۳$ q نتیجه می‌شود

$$A_۳(n, ۳) \geq \lceil ۳^n / \left(۱ + ۲\binom{n}{۱} + ۲^۲\binom{n}{۲} \right) \rceil = \lceil ۳^n/(۲n^۲ + ۱) \rceil.$$

۱۲.۶ برای $n = ۱$ داریم $H = (۱)$ یا $H = (-)$. برای $n = ۲$ داریم

$$H = \pm \begin{pmatrix} ۱ & ۱ \\ ۱ & - \end{pmatrix}, \pm \begin{pmatrix} ۱ & ۱ \\ - & ۱ \end{pmatrix}, \pm \begin{pmatrix} ۱ & - \\ ۱ & ۱ \end{pmatrix} \quad \text{یا} \quad \pm \begin{pmatrix} ۱ & - \\ - & - \end{pmatrix}.$$

۱۳.۶ چون درایه‌های H همگی ± ۱ هستند وضعیت مشابه‌ای برای H' برقرار است، و به‌سادگی می‌توان نشان داد که سطرهای متمایز H متعامد هستند.

۱۴.۶ ۱۱۱۱، ۰۰۰۰، ۱۰۱۰، ۰۱۰۱، ۱۱۰۰، ۰۰۱۱، ۱۰۰۱، ۰۱۱۰. اینها کد دوتایی کنترل تساوی $\mathcal{P}_۴$ را تشکیل می‌دهند که یک کد خطی است.

۱۵.۶ با اعمال لم ۲۴.۶ روی ماتریس هادامارد مرتبه ۴ در مثال ۲۶.۶، یک ماتریس هادامارد از مرتبه ۸ حاصل می‌شود

$$H' = \begin{pmatrix} ۱ & ۱ & ۱ & ۱ & ۱ & ۱ & ۱ & ۱ \\ ۱ & - & ۱ & - & ۱ & - & ۱ & - \\ ۱ & ۱ & - & - & ۱ & ۱ & - & - \\ ۱ & - & - & ۱ & ۱ & - & - & ۱ \\ ۱ & ۱ & ۱ & ۱ & - & - & - & - \\ ۱ & - & ۱ & - & - & ۱ & - & ۱ \\ ۱ & ۱ & - & - & - & - & ۱ & ۱ \\ ۱ & - & - & ۱ & - & ۱ & ۱ & - \end{pmatrix}$$

که ۱۶ کدکلمه زیر را ایجاد می‌کند

۱۱۱۱۱۱۱۱ ۰۰۰۰۰۰۰۰ ۱۰۱۰۱۰۱۰ ۰۱۰۱۰۱۰۱
۱۱۰۰۱۱۰۰ ۰۰۱۱۰۰۱۱ ۱۰۰۱۱۰۰۱ ۰۱۱۰۰۱۱۰
۱۱۱۱۰۰۰۰ ۰۰۰۰۱۱۱۱ ۱۰۱۰۰۱۰۱ ۰۱۰۱۱۰۱۰
۱۱۰۰۰۰۱۱ ۰۰۱۱۱۱۰۰ ۱۰۰۱۰۱۱۰ ۰۱۱۰۱۰۰۱

نرخ آن $\log_2 16)/8 = 1/2$ است. چون $d = 4$ پس $t = 1$؛ این کد $d - 1 = 3$ خطا را کشف می‌کند.

۱۶.۶ یک چندجمله‌ای $f(x)$ از درجه ۳ تحویل‌ناپذیر است اگر و فقط اگر عامل خطی نداشته باشد، یعنی ریشه نداشته باشد، پس $f(x) = x^3 + x + 1$ و $g(x) = x^3 + x^2 + 1$ تنها انتخاب‌های ممکن هستند. اگر α و β ریشه‌های f و g باشند، آنگاه $F = \{a\alpha^2 + b\alpha + c \mid a, b, c \in \mathbf{Z}_2\}$ و $F' = \{a\beta^2 + b\beta + c \mid a, b, c \in \mathbf{Z}_2\}$ میدان‌های از مرتبه ۸ بوده با این خاصیت که $\alpha^3 = \alpha + 1$ و $\beta^3 = \beta^2 + 1$. در این‌صورت $(\beta + 1)^3 = (\beta + 1) + 1$، و از این‌رو نگاشت

$$a\alpha^2 + b\alpha + c \longmapsto a(\beta + 1)^2 + b(\beta + 1) + c = a\beta^2 + b\beta + (a + b + c)$$

یک ایزومورفیسم از F به F' است.

۱۷.۶ اگر $f(x) = x^2 + 1$ یک ریشه α در $\mathbf{Z}_p$ داشته باشد، آنگاه $\alpha^2 = -1 \neq 1$ ولی $\alpha^4 = (-1)^2 = 1$، پس α در گروه ضربی $\mathbf{Z}_p^*$ از مرتبه ۴ است؛ پس ۴ بخش‌کننده $|\mathbf{Z}_p^*| = p - 1$ است، که غیرممکن است زیرا $p = 3 \bmod (4)$. پس ریشه α از f در $\mathbf{Z}_p$ نیست، بنابراین $F = \{a\alpha + b \mid a, b \in \mathbf{Z}_p\}$ یک میدان از مرتبه p^2 است با $\alpha^2 = -1$. با به‌کار بردن $x^3 - 1 = (x - 1)(x^2 + x + 1)$، یک بحث مشابه‌ای نشان می‌دهد اگر $p = 2 \bmod (3)$ آنگاه $x^2 + x + 1$ روی $\mathbf{Z}_p$ تحویل‌ناپذیر است، که در این حالت یک میدان $F = \{a\alpha + b \mid a, b \in \mathbf{Z}_p\}$ از مرتبهٔ p^2 با خاصیت $\alpha^2 + \alpha + 1 = 0$ وجود دارد.

۱۸.۶ چون تمامی کدکلمه‌ها در حداقل d موقعیت متفاوت هستند، حذف $d - 1$ سمبل منجر به ایجاد یک مجموعه از M کلمه متمایز به‌طول $n - d + 1$ روی F_q می‌شود. حداکثر q^{n-d+1} کلمه این‌چنینی وجود دارد، پس $M \leq q^{n-d+1}$. حال لگاریتم بگیرید. یک کد تکراری $\mathbf{R}_n$ این کران را با $M = q$ و $d = n$ به‌دست می‌آورد، همچنان‌که کد کنترل تساوی $\mathcal{P}_n$ با $M = q^{n-1}$ و $d = 2$ این کران را اختیار می‌کند.

۱۹.۶ از $۲۰۴۸ = ۲^{۱۱} = \binom{۲۳}{۰} + \binom{۲۳}{۱} + \binom{۲۳}{۲} + \binom{۲۳}{۳}$ نتیجه می‌شود
$۲^{۱۲}\left(\binom{۲۳}{۰} + \binom{۲۳}{۱} + \binom{۲۳}{۲} + \binom{۲۳}{۳}\right) = ۲^{۲۳}$;
این منجر به برقراری تساوی در کران کره‌چینی همینگ با $n = ۲۳$، $q = ۲$، $t = ۳$ و $M = ۲^{۱۲}$ می‌شود. مشابهاً از $۲۴۳ = ۳^۵ = ۲^۲.\binom{۱۱}{۲} + ۲.\binom{۱۱}{۱} + \binom{۱۱}{۰}$ وجود این تساوی با پارامترهای $n = ۱۱$، $q = ۳$، $t = ۲$، و $M = ۳^۶$ نتیجه می‌شود. این وجود کدهای خطی کامل با ابعاد ۱۲ و ۶ را ندا می‌دهد. اینها کدهای گلی هستند که در بخش ۵.۷ توصیف شدند. (با این حال تمرین ۱۶.۷ را ببینید.)

۲۰.۶ $\mathcal{C}_۱ \oplus \mathcal{C}_۲$ و $\mathcal{C}_۱ * \mathcal{C}_۲$ زیرمجموعه‌هایی از فضای برداری $۲n$ بعدی $\mathcal{V} \oplus \mathcal{V}$ هستند، از این‌رو کدهای به‌طول $۲n$ می‌باشند. تعداد $M_۱$ بردار $\mathbf{x}$ و $M_۲$ بردار $\mathbf{y}$ منجر به $M_۱M_۲$ بردار متمایز $(\mathbf{x},\mathbf{y})$ یا $(\mathbf{x},\mathbf{x}+\mathbf{y})$ می‌شوند، پس هر کد شامل $M_۱M_۲$ کدکلمه است. اعضای $(\mathbf{x},\mathbf{y})$ و $(\mathbf{x}',\mathbf{y}')$ از $\mathcal{C}_۱ \oplus \mathcal{C}_۲$ متمایز هستند اگر و فقط اگر $\mathbf{x} \neq \mathbf{x}'$ یا $\mathbf{y} \neq \mathbf{y}'$، که در این حالت

$$d((\mathbf{x},\mathbf{y}),(\mathbf{x}',\mathbf{y}')) = d(\mathbf{x},\mathbf{x}') + d(\mathbf{y},\mathbf{y}') \geq \min(d_۱, d_۲);$$

این کران با انتخاب $\mathbf{x} = \mathbf{x}'$ یا $\mathbf{y} = \mathbf{y}'$ و نزدیک کردن زوج متمایز دیگر تا حد ممکن به‌دست می‌آید، پس $d(\mathcal{C}_۱ \oplus \mathcal{C}_۲) = \min(d_۱, d_۲)$. در $\mathcal{C}_۱ * \mathcal{C}_۲$ اگر $\mathbf{x} = \mathbf{x}'$ و $\mathbf{y} \neq \mathbf{y}'$ آنگاه $d((\mathbf{x},\mathbf{x}+\mathbf{y}),(\mathbf{x}',\mathbf{x}'+\mathbf{y}')) = d(\mathbf{y},\mathbf{y}')$ دارای $d_۲$ بوده و اگر $\mathbf{x} \neq \mathbf{x}'$ و $\mathbf{y} = \mathbf{y}'$ آنگاه $d((\mathbf{x},\mathbf{x}+\mathbf{y}),(\mathbf{x}',\mathbf{x}'+\mathbf{y}')) = ۲d(\mathbf{x},\mathbf{x}')$ دارای کمترین مقدار $۲d_۱$ خواهد بود؛ اگر $\mathbf{x} \neq \mathbf{x}'$ و $\mathbf{y} \neq \mathbf{y}'$ آنگاه $d(\mathbf{x},\mathbf{x}') \geq d_۱$ و $d(\mathbf{x}+\mathbf{y},\mathbf{x}'+\mathbf{y}') \geq |d_۱ - d_۲|$، پس

$$d((\mathbf{x},\mathbf{x}+\mathbf{y}),(\mathbf{x}',\mathbf{x}'+\mathbf{y}')) \geq d_۱ + |d_۱ - d_۲| \geq d_۲ \geq \min(۲d_۱, d_۲)$$

و بنابراین $d(\mathcal{C}_۱ * \mathcal{C}_۲) = \min(۲d_۱, d_۲)$. اگر $\mathcal{C}_i$ ها خطی باشند آنگاه $\mathcal{C}_۱ \oplus \mathcal{C}_۲$ و $\mathcal{C}_۱ * \mathcal{C}_۲$ زیرفضاهای خطی $\mathcal{V} \oplus \mathcal{V}$ هستند (بسته بودن به آسانی بررسی می‌شود)، و بنابراین کدهای خطی هستند؛ بعد آنها برابر $\log_q M_۱M_۲ = \log_q M_۱ + \log_q M_۲ = k_۱ + k_۲$ است.

۲۱.۶ اگر a_j (یعنی رقم j ام) به $b_j \neq a_j$ تغییر یابد آنگاه $\sum_{i=۱}^{۱۰} ia_i (\equiv ۰ \bmod (۱۱))$ با

$$\sum_{i \neq j} ia_i + jb_j = \sum_i ia_i + j(b_j - a_j) \equiv ۰ + j(b_j - a_j) \equiv j(b_j - a_j) \not\equiv ۰$$

جایگزین می‌شود که ناصفری این مقدار به دلیل $j, b_j - a_j \not\equiv ۰$ است، بنابراین خطا کشف می‌شود. مشابهاً، اگر a_j و a_k جابه‌جا شوند، که $a_j \neq a_k$، آنگاه $\sum_i ia_i$ با

$$\sum_i ia_i + j(a_k - a_j) + k(a_j - a_k) \equiv ۰ + (j-k)(a_k - a_j) \not\equiv ۰$$

جایگزین می‌شود، و بنابراین خطا کشف می‌شود. در هر حالت، اول بودن ۱۱ مهم است، به‌قسمی‌که $x, y \not\equiv 0$ ایجاب می‌کند $xy \not\equiv 0$ (که این برای هنگ مرکب غلط است). برای ۷ - ۷۶۱۹۷ - ۵۴۰ - ۳ داریم $a_1 + 2a_2 + \cdots + 10a_{10} = 308 = 0 \bmod 11$، پس این یک ISBN معتبر است (در واقع شماره متعلق به کتاب نظریه مقدماتی اعداد، به‌وسیله جانز و جانز، است)؛ تفاوت دومی و سومی با این در یک جایگشت و یک خطا است و بنابراین ISBN نیستند.

۱.۷ با اضافه نمودن یک ستون به G به‌قسمی‌که مجموع درایه‌های هر سطر صفر باشد $\overline{G}$ را تشکیل دهید. $\overline{H}$ را از روی H با اضافه نمودن یک ستون با $c = n - k$ درایه ۰ و سپس یک سطر با $n + 1$ درایه ۱ بسازید.

۲.۷ یک ماتریس مولد G برای $\mathcal{C}_1 + \mathcal{C}_2$ با الحاق سطرهای G_2 به G_1 و سپس حذف سطرهای وابسته با استفاده از عملیات سطری مقدماتی تشکیل دهید. یک روند مشابه روی سطرهای H_1 و H_2 منجر به ساخت ماتریس کنترل تساوی H برای $\mathcal{C}_1 \cap \mathcal{C}_2$ می‌شود.

۳.۷ هر سطر از G_1 یک ترکیب خطی از سطرهای G_2 است، پس $\mathcal{C}_1 \subseteq \mathcal{C}_2$؛ چون $\dim \mathcal{C}_1 = \dim \mathcal{C}_2$ پس $\mathcal{C}_1 = \mathcal{C}_2$. معادلاً $G_2 H^T = O$ که H ماتریس کنترل تساوی برای $\mathcal{C}_1 = \mathcal{H}_3$ در مثال ۱۳.۷ است، پس $\mathcal{C}_1 \subseteq \mathcal{C}_2$؛ از مقایسه ابعاد تساوی نتیجه می‌شود.

۴.۷ چون $\mathcal{H}_n$ کد کامل ۱ خطا تصحیح کننده است، کدگشایی نزدیک‌ترین همسایه خطاهای به وزن حداکثر ۱ را درست می‌کند ولی بقیه را خیر؛ احتمال هیچ خطا برابر P^n و احتمال بروز یک خطا در یک موقعیت مفروض برابر $P^{n-1}Q$ است، پس $Pr_C = P^n + nP^{n-1}Q$ و $Pr_E = 1 - Pr_C = 1 - P^n - nP^{n-1}Q$. اگر $P < 1$ آنگاه P^n و nP^{n-1} با افزایش n به ۰ میل می‌کنند پس $Pr_E \longrightarrow 1$. اگر $P = 1$ آنگاه برای هر n داریم $Pr_E = 0$.

۵.۷ $\mathbf{u}H^T = 0$، پس $\mathbf{u} \in \mathcal{H}_3$. مشخصه $\mathbf{v}$ برابر $\mathbf{s} = \mathbf{v}H^T = 101 = \mathbf{c}_2^T$ است، پس خطا در موقعیت دوم بوده و بنابراین $\Delta(\mathbf{v}) = \mathbf{v} - \mathbf{e}_2 = 1100110$؛ این $\mathbf{u}$ است و از این‌رو کدگشایی درست است. با این حال، $\mathbf{v}'$ دارای مشخصه $\mathbf{s}' = \mathbf{v}'H^T = 110 = \mathbf{c}_4^T$ بوده و بنابراین $\Delta(\mathbf{v}') = \mathbf{v}' - \mathbf{e}_3 = 0010110 \neq \mathbf{u}$، پس کدگشایی نادرست است. این بدین جهت است که $\mathbf{v}'$ حاوی دو خطا و $\mathbf{v}$ حاوی یک خطا بوده و $\mathcal{H}_3$ تنها یک خطا را تصحیح می‌کند.

۶.۷ مشخصه $\mathbf{s} = \mathbf{v}H^T = 010$ نمایش دوتایی ۲ است و اشاره به وجود خطا در موقعیت دوم دارد؛ پس $\Delta(\mathbf{v}) = \mathbf{v} - \mathbf{e}_2 = 0111100 = \mathbf{u}$ و از این‌رو کدگشایی درست است.

۷.۷ هیچ برداری نمی‌تواند مضربی از دیگری باشد، پس آنها باید زیرفضاهای یک بعدی متمایزی را تولید کنند. تعداد زیرفضاهای این چنینی در $\mathcal{W}$ برابر $n = (q^c - ۱)/(q - ۱)$ است، پس این بیشترین تعداد بردار است. اگر آنها ستون‌های H باشند آنگاه کد متناظر C به‌طول n و بعد $k = n - c$ خواهد بود. هیچ دو ستونی وابسته خطی نیستند ولی سه ستون وابسته وجود دارد (جمع هر دو برداری مضربی از یک بردار سوم است)، پس $d = ۳$ و $t = ۱$. سپس $\sum_{i=۰}^{t} \binom{n}{i}(q-۱)^i = ۱ + n(q-۱) = q^c = q^{n-k}$، پس C کامل است.

۸.۷ هر $s \in S$ در $\binom{۲۳}{۴}$ زیرمجموعه ۵ عضوی از S قرار دارد، که هریک از آنها در یک قالب منحصربه‌فرد قرار دارد؛ به‌عکس، هر قالب شامل s حاوی $\binom{۷}{۴}$ زیرمجموعه ۵ عضوی حاوی s است، بنابراین s در $\binom{۷}{۴}/\binom{۲۳}{۴} = ۲۵۳$ قالب قرار دارد. مشابهاً تعداد قالب‌های حاوی یک زوج، یک سه‌تایی و یک چهارتایی برابر $\binom{۶}{۴}/\binom{۲۲}{۳} = ۷۷$، $\binom{۵}{۲}/\binom{۲۱}{۲} = ۲۱$، و $\binom{۴}{۱}/\binom{۲۰}{۱} = ۵$ است.

۹.۷ ماتریس مولد داده شده ماتریس زیر است

$$\begin{pmatrix} ۰ & ۱ & ۱ & ۰ & ۱ & ۱ \\ ۱ & ۰ & ۱ & ۱ & ۰ & ۱ \\ ۱ & ۱ & ۱ & ۰ & ۰ & ۰ \end{pmatrix}.$$

با اضافه نمودن سطر ۳ به سطرهای ۱و۲، و سپس اضافه نمودن سطرهای جدید ۱ و ۲ به سطر ۳، ماتریس G و بنابراین ماتریس H به‌دست می‌آیند

$$G = \begin{pmatrix} ۱ & ۰ & ۰ & ۰ & ۱ & ۱ \\ ۰ & ۱ & ۰ & ۱ & ۰ & ۱ \\ ۰ & ۰ & ۱ & ۱ & ۱ & ۰ \end{pmatrix}, \qquad H = \begin{pmatrix} ۰ & ۱ & ۱ & ۱ & ۰ & ۰ \\ ۱ & ۰ & ۱ & ۰ & ۱ & ۰ \\ ۱ & ۱ & ۰ & ۰ & ۰ & ۱ \end{pmatrix}.$$

۱۱۰ به‌عنوان $\mathbf{c} = ۱۱۰$ $G = ۱۱۰۱۱۰$ کدگذاری می‌شود و $\mathbf{c}H^T = ۰۰۰ = \mathbf{0}$. چون $n = ۶$ و $k = ۳$، نرخ برابر $R = k/n = \frac{۱}{۲}$ است. می‌نیمم فاصله d برابر کمترین تعداد ستون وابسته خطی در H است و این عدد برابر ۳ است (ستون‌های ۱ و ۲ و ۳). جدول مشخصه

$\mathbf{v}_i =$	۰۰۰۰۰۰	۱۰۰۰۰۰	۰۱۰۰۰۰	۰۰۱۰۰۰	۰۰۰۱۰۰	۰۰۰۰۱۰	۰۰۰۰۰۱	۱۰۰۱۰۰
$\mathbf{s}_i =$	۰۰۰	۰۱۱	۱۰۱	۱۱۰	۱۰۰	۰۱۰	۰۰۱	۱۱۱

تمامی خطاهای به وزن یک را به‌انضمام بردار خطای ۱۰۰۱۰۰ تصحیح می‌کند، پس اگر هر سمبل دارای احتمال P و Q به‌ترتیب برای ارسال درست و نادرست باشد آنگاه

$$Pr_E = ۱ - (P^۶ + ۶P^۵Q + P^۴Q^۲).$$

۱۰.۷ هر $\mathcal{C}$ دارای یک مبنای مرتب $\mathbf{u}_1,\dots,\mathbf{u}_k$ است. تعداد q^n-1 انتخاب برای $\mathbf{u}_1\in\mathcal{V}$ وجود دارد، سپس q^n-q انتخاب برای $\mathbf{u}_2$ (کنار گذاشتن مضارب $\mathbf{u}_1$)، $\dots$، q^n-q^{k-1} انتخاب برای $\mathbf{u}_k$، از این‌رو تعداد یک چنین مبنایی برابر $(q^n-1)\dots(q^n-q^{k-1})$ است. مشابهاً هر $\mathcal{C}$ دارای $(q^k-1)\dots(q^k-q^{k-1})$ مبنای مرتب می‌باشد، پس تعداد کدهای $\mathcal{C}$ برابر است با:

$$(q^n-1)\dots(q^n-q^{k-1})/(q^k-1)\dots(q^k-q^{k-1}).$$

۱۱.۷ $L_1\cap L_2$ یک نقطه p است که روی یک خط منحصربه‌فرد سومی چون L_3 قرار دارد؛ سه مجموعه $L_i\backslash\{p\}$ مجموعه $S\backslash\{p\}$ را افراز می‌کنند، پس $L_1+L_2=(L_1\backslash\{p\})\cup(L_2\backslash\{p\})=(S\backslash\{p\})\backslash(L_3\backslash\{p\})=S\backslash L_3$ که متمم L_3 یعنی $\overline{L}_3$ است. پس زیرفضای تولید شده با این خطوط حاوی هفت خط و متمم آنها $\overline{L}$ به همراه $L+L=\emptyset$ و $L+\overline{L}=S$ است. این مجموعه متشکل از شانزده زیرمجموعه از S تحت عمل جمع بسته است (به‌عنوان مثال $L_3+\overline{L}_2=L_1+\overline{L}_2$ و $\overline{L}_1+\overline{L}_2=\overline{L}_3$)، و بنابراین برابر $\mathcal{C}$ است. از این‌رو $\mathcal{C}$ یک کد خطی دوتایی به‌طول $n=7$ و بعد $k=\log_2 16=4$ است. کدکلمات ناصفر L، $\overline{L}$ و S به وزن $|L|=3$، $|\overline{L}|=4$ و $|S|=7$ می‌باشند، پس $\mathcal{C}$ دارای $d=3$ و $t=1$ است. در بخش ۴.۷ نشان دادیم که هر دو $[7,4]$–کد خطی دوتایی ۱ خطا تصحیح کننده معادل هستند، از این‌رو $\mathcal{C}$ معادل $\mathcal{H}_7$ است.

۱۲.۷ هر دو نقطه‌ای محمل یک بردار $\mathbf{v}$ به وزن ۲ را تشکیل می‌دهند؛ چون $\mathcal{C}$ کامل است و $t=1$، $\mathbf{v}$ به فاصله ۱ از یک کدکلمه منحصربه‌فرد $\mathbf{u}$ به وزن ۲ است که محمل آن قالب منحصربه‌فرد شامل دو نقطه مفروض است. بنابراین کدکلمات به وزن ۳ قالب‌های یک سیستم اشتاینر $S(2,3,n)$ هستند. در $\mathcal{H}_n$، مختصات $i=1,\dots,n=2^c-1$، که به‌صورت دوتایی نوشته می‌شوند تا ستون‌های ماتریس کنترل تساوی H را تشکیل دهند، متشکل از بردارهای ناصفر در F_2^c می‌باشند، پس متناظر با نقاط در $PG(c-1,2)$ می‌شوند؛ کدکلمات به وزن ۳ متناظر با روابط $\mathbf{c}_i+\mathbf{c}_j+\mathbf{c}_k=\mathbf{0}$ بین ستون‌های H و بنابراین متناظر با خطوط $\{\mathbf{c}_i,\mathbf{c}_j,\mathbf{c}_k\}$ از $PG(c-1,2)$ می‌باشند.

۱۳.۷ جایگشت همانی $\mathcal{C}$ را به خودش می‌نگارد، و اگر g و h چنین باشند آنگاه gh و g^{-1} نیز $\mathcal{C}$ را به خودش تصویر می‌کنند؛ بنابراین $\mathrm{Aut}(C)$ یک زیرگروه S_n است. $\mathcal{R}_n$ و $\mathcal{P}_n$ هر دو تحت هر جایگشتی پایدار هستند و بنابراین گروه اتومورفیسم آنها S_n است. کد $\mathcal{R}_2\oplus\mathcal{R}_2=\{0000,1100,0011,1111\}$ دارای هشت اتومورفیسم (۱۲)، (۳۴)، (۳۴)(۱۲)، (۲۴)(۱۳)، (۲۳)(۱۴)، (۱۳۲۴)، (۱۴۲۳) و نگاشت همانی (گروه دو سطحی) می‌باشد. کدهای معادل با $\mathcal{C}$ آنهایی هستند که از اعمال یک جایگشت روی مختصات $\mathcal{C}$ به‌دست می‌آیند؛ دو جایگشت کد یکسان تولید می‌کنند

اگر و فقط اگر آنها در یک هم‌دسته از $\mathrm{Aut}(\mathcal{C})$ در S_n قرار گیرند. پس تعداد کدهای معادل برابر تعداد هم‌دسته‌های $\mathrm{Aut}(\mathcal{C})$ یعنی $|S_n|/|\mathrm{Aut}(\mathcal{C})| = n!/|\mathrm{Aut}(\mathcal{C})|$ می‌باشد. اگر $\mathcal{C} = \mathcal{R}_2 \oplus \mathcal{R}_2$ آنگاه $4!/8 = 3$ کد معادل وجود دارد که عبارت هستند از $\mathcal{C}$ و $\{0000, 1010, 0101, 1111\}$ و $\{0000, 0110, 1001, 1111\}$.

۱۴.۷ بنابر تمرین‌های ۱۱.۷ و ۱۲.۷، هر اتومورفیسم از $PG(2,2)$ یک اتومورفیسم از $\mathcal{H}_7$ را ایجاد نموده و به‌عکس، پس گروه اتومورفیسم آنها ایزومورف هستند. اتومورفیسم‌های $PG(2,2)$ از اتومورفیسم‌های فضای برداری F_2^3 نتیجه می‌شوند و اینها گروه خطی $GL(3,2)$ متشکل از ماتریس‌های 3×3 معکوس‌پذیر روی F_2 را می‌سازند؛ تنها ماتریس همانی منجر به اتومورفیسم همانی در $PG(2,2)$ می‌شود، بنابراین $\mathrm{Aut}(\mathcal{H}_7) \cong Aut\ (PG(2,2)) \cong GL(3,2)$. تعداد $7 = 1 - 2^3$ امکان برای اولین سطر از یک ماتریس در $GL(3,2)$ وجود دارد؛ وقتی این سطر اول انتخاب شده است، تعداد $6 = 2 - 2^3$ امکان برای سطر دوم وجود دارد، و سپس $4 = 2^2 - 2^3$ امکان برای سطر سوم، پس $|\mathrm{Aut}(\mathcal{H}_7)| = |GL(3,2)| = 7.6.4 = 168$. بنابر تمرین ۱۳.۷ تعداد $168/7! = 30$ کد معادل با $\mathcal{H}_7$ وجود دارد. مشابهاً اگر $n = 2^c - 1$ آنگاه $\mathrm{Aut}(PG(c-1,2)) \cong GL(c,2)$ و $GL(c,2)$ از مرتبه $(2^c - 1)(2^c - 2)(2^c - 2^2)\dots(2^c - 2^{c-1})$ است، پس تعداد $n!/(2^c - 1)(2^c - 2)(2^c - 2^2)\dots(2^c - 2^{c-1})$ کد معادل وجود دارد.

۱۵.۷ هر $t+1$ نقطه‌ای محمل یک بردار به وزن $t+1$ است؛ چون $\mathcal{C}$ کامل است این بردار به فاصله t از یک کدکلمه منحصربه‌فرد به وزن $d = 2t+1$ است که محمل آن قالب منحصربه‌فردی است که شامل $t+1$ نقطه مفروض می‌باشد. پس یک سیستم اشتاینر $S(t+1, d, n)$ داریم (تمرین ۱۲.۷ را برای حالت $t = 1$ ببینید). تعداد قالب‌ها برابر $\binom{n}{t+1}/\binom{d}{t+1}$ است، پس $\binom{d}{t+1}$ بخش‌کننده $\binom{n}{t+1}$ است. با حذف i نقطه، که $1 \le i \le t$، سیستم اشتاینر $S(t+1-i, d-i, n-i)$ حاصل می‌شود، بنابراین با دلیل مشابه $\binom{d-i}{t+1-i}$ بخش‌کننده $\binom{n-i}{t+1-i}$ است.

۱۶.۷ $2^{12} = 4096 = \binom{90}{2} + 90 + 1$، پس پارامترهای $q = 2$، $n = 90$، $t = 2$، $M = 2^{78}$ منجر به تساوی در کران کره‌چینی همینگ می‌شوند که اشاره‌ای به امکان وجود یک کد کامل دوتایی ۲ خطا تصحیح کننده به طول ۹۰ دارد. ولی اگر این کد وجود داشته باشد آنگاه با قرار دادن $d = 2t+1 = 5$ و انتخاب $i = 2$ در تمرین ۱۵.۷ می‌بینیم که ۳ بخش‌کننده ۸۸ است که غلط است. (انتخاب $i = 1$ منجر به یک تناقض با آشکاری کمتری می‌شود.)

۱۷.۷ هر موقعیت مختصاتی i به هر یک از طرفین معادله کمکی برابر ۱ یا ۰ دارد. دوازده سطر G به‌طول ۲۴ بوده و مستقل هستند، بنابراین آنها یک $[۲۴, ۱۲]$–کد خطی $\mathcal{C}$ دوتایی تولید می‌کنند. هر رأس بیست‌وجهی مجاور به ۵ رأس دیگر است، پس تعداد ۱ های موجود در هر سطر $\mathbf{r}$ از G زوج $(۸ = (۵ - ۱۲) + ۱)$ بوده و بنابراین $\mathbf{r}.\mathbf{r} = ۰$؛ مشابهاً هر دو رأس متمایزی دارای یک تعداد زوجی از رئوس مشترک غیر همسایه می‌باشند، و بنابراین سطرهای G دوبه‌دو متعامد بوده و از این‌رو $\mathcal{C} \subseteq \mathcal{C}^{\perp}$؛ چون $\dim \mathcal{C} = \dim(\mathcal{C}^{\perp})$ داریم $\mathcal{C} = \mathcal{C}^{\perp}$. چون P دوتایی و متقارن است $(P|I) = (-P^T|I)$؛ این یک ماتریس کنترل برای $\mathcal{C}$ است و بنابراین یک ماتریس مولد برای $\mathcal{C} = \mathcal{C}^{\perp}$ می‌باشد. هر سطر G دارای وزنی بخش‌پذیر بر ۴ است، و بنابر اولین نتیجه این خاصیت برای حاصل‌جمع دو کدکلمه نیز حفظ می‌شود زیرا از خود—دوگانی نتیجه می‌شود که $c(\mathbf{u}, \mathbf{v})$ همیشه زوج است. اگر $\mathbf{u} \in \mathcal{C}$ دارای x و y درایه ۱ در نیمه اول و نیمه دوم باشد به‌قسمی‌که $x + y = wt(\mathbf{v})$، آنگاه $\mathbf{u}$ مجموعی از x سطر از G و y سطر از G' است؛ اگر $wt(\mathbf{u}) = ۴$ آنگاه حاصل‌جمعی از حداکثر ۲ سطر از G (یا معادلاً از G') دارای وزن ۴ است که با دقت روی G غلط بودن آن مشخص می‌شود. پس $\mathcal{C}$ دارای می‌نیمم فاصله ۸ است و بنابراین $[۲۳, ۱۲]$– کد خطی دوتایی $\mathcal{C}^{\circ}$ دارای $d = ۷$ و از این‌رو $t = ۳$ می‌باشد. چون $\sum_{i=۰}^{۳} \binom{۲۳}{i} = ۲^{۲۳-۱۲}$ پس $\mathcal{C}^{\circ}$ کامل است.

۱۸.۷ شش سطر G مستقل بوده و به‌طول ۱۲ می‌باشند، بنابراین آنها یک $[۱۲, ۶]$– کد خطی سه‌تایی تولید می‌کنند. با دقت روی G ملاحظه می‌شود که سطرها دوبه‌دو متعامد هستند و بنابراین $\mathcal{C} \subseteq \mathcal{C}^{\perp}$؛ با مقایسه ابعاد به $\mathcal{C} = \mathcal{C}^{\perp}$ می‌رسیم. هر $\mathbf{u} \in \mathcal{C}$ در $\sum_i u_i^۲ = ۰$ (در $F_۳$) صدق کرده و از این‌رو وزن آن مضربی از ۳ است؛ بنابراین کافی است نشان دهیم $wt(\mathbf{u}) \neq ۳$، و این با درنظر گرفتن ترکیبات خطی مختلف سطرهای G مشخص می‌شود، پس $\mathcal{C}$ دارای می‌نیمم فاصله ۶ است. بنابراین $\mathcal{C}^{\circ}$ یک $[۱۱, ۶]$– کد سه‌تایی خطی با می‌نیمم فاصله ۵ است؛ این کد ۲ خطا را تصحیح کرده و چون $\sum_{i=۰}^{۲} \binom{۱۱}{i} ۲^i = ۳^{۱۱-۶}$ پس یک کد کامل است.

۱۹.۷ خواص اساسی $\mathcal{RM}(r, m)$ از تمرین ۲۰.۶ (با به‌کار بردن تعریف استقرایی این کدها) نتیجه می‌شوند. برای نمونه، $\mathcal{RM}(۰, m)$ و $\mathcal{RM}(m, m)$ خطی و دوتایی هستند. خواص با $*$ حفظ می شوند، بنابراین هر کد رید — مولر خطی و دوتایی است. چون $*$ منجر به دوبرابر شدن طول می‌شود، $\mathcal{RM}(r, m)$ دارای طول $n = ۲^m$ است. اگر $\mathcal{RM}(r, m-۱)$ و $\mathcal{RM}(r-۱, m-۱)$ به‌ترتیب دارای می‌نیمم فاصله $d_۱ = ۲^{m-۱-r}$ و $d_۲ = ۲^{m-r}$ باشند آنگاه $\mathcal{RM}(r, m)$ دارای می‌نیمم فاصله $d = \min(۲d_۱, d_۲) = ۲^{m-r}$

است. اگر $\mathcal{RM}(r,m-1)$ و $\mathcal{RM}(r-1,m-1)$ به‌ترتیب با ابعاد $k_1=\sum_{i=0}^{r}\binom{m-1}{i}$ و $k_2=\sum_{i=0}^{r-1}\binom{m-1}{i}$ باشند آنگاه $\mathcal{RM}(r,m)$ دارای بعد k_1+k_2 است:

$$\begin{aligned}k_1+k_2 &= \binom{m-1}{0}+\sum_{i=1}^{r}\binom{m-1}{i}+\sum_{i=0}^{r-1}\binom{m-1}{i}\\ &= \binom{m-1}{0}+\sum_{i=1}^{r}\binom{m-1}{i}+\sum_{i=1}^{r}\binom{m-1}{i-1}\\ &= \binom{m}{0}+\sum_{i=1}^{r}\binom{m}{i}=\sum_{i=0}^{r}\binom{m}{i}.\end{aligned}$$

$$\begin{aligned}\mathcal{RM}(1,2) &= \mathcal{RM}(1,1)*\mathcal{RM}(0,1)=\{00,01,10,11\}*\{00,11\}\\ &= \{0000,0011,0101,0110,1010,1001,1111,1100\}.\end{aligned}$$

$$\begin{aligned}\mathcal{RM}(1,3) &= \mathcal{RM}(1,2)*\mathcal{RM}(0,2)=\mathcal{RM}(1,2)*\{0000,1111\}\\ &= \{00000000,00110011,01010101,01100110,\\ &\quad 10101010,10011001,11111111,11001100,\\ &\quad 00001111,00111100,01011010,01101001,\\ &\quad 10100101,10010110,11110000,11000011\}.\end{aligned}$$

چون $\dim\mathcal{RM}(1,3)=4$، یک مبنا متشکل از ۴ کدکلمه مستقل مانند ۱۰۰۱۰۱۱۰، ۰۱۰۱۰۱۰۱، ۰۰۱۱۰۰۱۱، ۰۰۰۰۱۱۱۱ می‌باشد که این ماتریس مولد G را می‌سازد

$$G=\begin{pmatrix}1&0&0&1&0&1&1&0\\0&1&0&1&0&1&0&1\\0&0&1&1&0&0&1&1\\0&0&0&0&1&1&1&1\end{pmatrix}.$$

این ماتریس مولد به‌فرم منظم نیست، ولی جابه‌جایی ستون‌های ۴ و ۵ ماتریس مولد $G'=(I_4|P)$ را تولید می‌کند که منظم است و از آن‌جا یک ماتریس کنترل تساوی $H'=(-P^T|I_4)$ برای یک کد معادل حاصل می‌شود. جابه‌جایی ستون‌های ۴ و ۵ در ماتریس H' ماتریس کنترل تساوی H را برای $\mathcal{RM}(1,3)$ می‌سازد:

$$H=\begin{pmatrix}1&1&1&1&0&0&0&0\\1&1&0&0&1&1&0&0\\1&0&1&0&1&0&1&0\\0&1&1&0&1&0&0&1\end{pmatrix}$$

هیچ مجموعه‌ای متشکل از یک، دو، و یا سه ستون از H وابسته خطی نیست، ولی $c_1 + c_2 + c_7 + c_8 = 0$، بنابراین $\mathcal{RM}(1,3)$ دارای می‌نیمم فاصله $d = 4$ است.

کتاب‌نامه

[An74] I. Anderson, First Course in Combinatorial Mathematics, Oxford University Press, Oxford, 1974.

[As65] R. Ash, Information Theory, Wiley, New York, 1965.

[Ba63] G. Bandyopadhyay, A simple proof of the decipherability criterion of Sardinas and Patterson, Information and Control 6 (1963), 331-336.

[Be68] E. R. Berlekamp, Algebraic Coding Theory, McGraw-Hill, New York, 1968.

[Be74] E. R. Berlekamp (ed.), Key Papers in the Development of Coding Theory, IEEE Press, New York, 1974.

[BP85] J. Berstel and D. Perrin, Theory of Codes, Academic Press, Orlando, 1985.

[Bi65] P. Billingsley, Ergodic Theory and Information, Wiley, New York, 1965.

[BM75] I. F. Blake and R. C. Mullin, The Mathematical Theory of Coding, Academic Press, New York, 1975.

[BM76] I. F. Blake and R. C. Mullin, Introduction to Algebraic and Combinatorial Coding Theory, Academic Press, New York, 1976.

[BR98] T. S. Blyth and E. F. Robertson, Basic Linear Algebra, Springer Undergraduate Mathematics Series, Springer, London, 1998.

[Br56] L. Brillouin, Science and Information Theory, Academic Press, New York, 1956.

[CL91] P. J. Cameron and J. H. van Lint, Designs, Graphs, Codes and their Links, LMS Student Texts 22, Cambridge University Press, Cambridge, 1991.

[Ch85] W. G. Chambers, Basics of Communications and Coding, Oxford University Press, Oxford, 1985.

[CS92] J. H. Conway and N. J. A. Sloane, Sphere Packings, Lattices and Groups (2nd ed.), Springer-Verlag, New York, 1992.

[De74] N. Deo, Graph Theory with Applications to Engineering and Computer Science, Prentice-Hall, Englewood Cliffs, 1974.

[Ev63] S. Even, Tests for unique decipherability, IEEE Trans. Information Theory IT-9 (1963),109-112.

[Ev79] S. Even, Graph Algorithms, Pitman, London, 1979.

[Fe50] W. Feller, Introduction to Probability Theory and its Applications, I, Wiley, New York, 1950.

[Fi83] E. Fisher, Intermediate Real Analysis, Springer- Verlag, New York, 1983.

[Gi52] E. N. Gilbert, A comparison of signalling alphabets, Bell System Tech. J. 31 (1952),504-522.

[GM59] E. N. Gilbert and E. F. Moore, Variable-length binary encodings, Bell System Tech. J. 38 (1959),933-967.

[Go49] M. J. E. Golay, Notes on digital coding, Proc. IEEE 37 (1949),657.

[Go80] S. W. Golomb, Sources which maximise the choice of a Huffman coding tree, Information and Control 45 (1980), 263-272.

[Go88] V. D. Goppa, Geometry and Codes, Kluwer, Dordrecht, 1988.

[Ha67] M. Hall, Jr, Combinatorial Theory, Blaisdell, Waltham Mass.,1967.

[Ha48] R. W. Hamming, Single error-correcting codes - Case 20878, Memorandum 48-110-52, Bell Telephone Laboratories, 1948.

[Ha50] R. W. Hamming, Error detecting and error correcting codes, Bell System Tech. J. 29 (1950), 147-160.

[Hi86] R. Hill, A First Course in Coding Theory, Oxford University Press, Oxford, 1986.

[Hu52] D. A. Huffman, A method for the construction of minimum redundancy codes, Proc. IRE 40 (1952), 1098-1101.

[Jo79] D. S. Jones, Elementary Information Theory, Oxford University Press, Oxford, 1979.

[Ka61] J. Karush, A simple proof of an inequality of McMillan, IRE Trans. Information Theory IT-7 (1961),118.

[Ke56] J. L. Kelley, Jr, A-new interpretation of information rate, Bell System Tech. J. 35 (1956), 917-926.

[KR83] K. H. Kim and F. W. Roush, Applied Abstract Algebra, Ellis Horwood, Chichester, 1983.

[Kn73] D. E. Knuth, The Art of Computer Programming, vol. I: Fundamental Algorithms, Addison-Wesley, Reading Mass., 1973.

[Kr49] L. G. Kraft, A device for quantizing, grouping, and coding amplitude modulated pulses, M. S. thesis, Electrical Engineering Department, MIT, 1949.

[La83] S. Lang, Undergraduate Analysis, Springer- Verlag, New York, 1983.

[Le64] V. I. Levenshtein, Some properties of coding and self adjusting automata for decoding messages, Problemy Kibernеticki 11 (1964),63-121. (Russian)

[Li82] J. H. van Lint, Introduction to Coding Theory, Springer - Verlag, NewYork, 1982.

[LG88] J. H. van Lint and G. H. van der Geer, Introduction to Coding Theory and Algebraic Geometry, DMV Seminar 12, Birkhauser, Basel, 1988.

[McE77] R. J. McEliece, The Theory of Information and Coding, Encyclopedia of Mathematics and its Applications 3, Addison-Wesley, Reading Mass., 1977.

[McM56] B. McMillan, Two inequalities implied by unique decipherability, IRE Trans. Information Theory IT-2 (1956), 115-116.

[MS77] F. J. MacWilliams and N. J. A. Sloane, The Theory of Error Correcting Codes, North-Holland, Amsterdam, 1977.

[Mu53] S. Muroga, On the capacity of a discrete channel, J. Phys. Soc. Japan 8 (1953), 484-494.

[Pl82] V. Pless, Introduction to the Theory of Error-Correcting Codes, Wiley, New York, 1982.

[PH98] V. S. Pless and W. Huffman (eds), Handbook of Coding Theory (2vols), Elsevier, Amsterdam, 1998.

[Pr92] O. Pretzel, Error-Correcting Codes and Finite Fields, Oxford University Press, Oxford, 1992.

[Re61] F. M. Reza, An Introduction to Information Theory, McGraw-Hill,New York, 1961.

[Ri67] J. A. Riley, The Sardinas-Patterson and Levenshtein theorems, Information and Control 10 (1967), 120-136.

[SP53] A. A. Sardinas and C. W. Patterson, A necessary and sufficient condition for the unique decomposition of coded messages, IRE. Internat. Conv. Rec. 8 (1953), 104-108.

[Sc64] E. S. Schwartz, An optimum encoding with minimum longest code and total number of digits, Information and Control 7 (1964),37-44.

[Sh48] C. E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27 (1948), 379-423, 623-656.

[SW63] C. E. Shannon and W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana, 1963.

[Se67] D. A. R. Seeley, A short note on Bandyopadhyay's proof of the decipherability criterion of Sardinas and Patterson, Information and Control 10 (1967),104-106.

[Si64] R. C. Singleton, Maximum distance q-nary codes, IEEE Thans. Information Theory IT-10 (1964), 116-118.

[Sl74] D. Slepian (ed.), Key Papers in the Development of Information Theory, IEEE Press, New York, 1974.

[St93] H. Stichtenoth, Algebraic Function Fields and Codes, Springer - Verlag,Berlin, 1993.

[ST81] M. N. S. Swamy and K. Thulasiraman, Graphs, Networks and Algorithms, Wiley, New York, 1981.

[Th83] T. M. Thompson, From Error-Correcting Codes through Sphere Packings to Simple Groups, Carus Mathematical Monographs 21,Math. Assoc. of America, 1983.

[Va57] R. R. Varshamov, Estimate of the number of signals in error correcting codes, Dokl. Akad. Nauk SSSR 117 (1957),739-741. (Russian)

[We88] D. Welsh, Codes and Cryptography, Oxford University Press, Oxford,1988.

[Zi59] S. Zimmerman, An optimal search procedure, Amer. Math. Monthly 66 (1959), 690-693.

لیست سمبل‌ها و اختصارها

سمبل ■ برای مشخص نمودن پایان یک اثبات به‌کار می‌رود. سمبل‌های زیر که معمولاً در ریاضیات به‌کار می‌روند بدون توضیح بیشتر مورد استفاده قرار می‌گیرند.

$\mathbf{C}$	مجموعه اعداد مختلط
$\mathbf{R}$	مجموعه اعداد حقیقی
$\mathbf{Q}$	مجموعه اعداد گویا
$\mathbf{Z}$	مجموعه اعداد صحیح
$\mathbf{N}$	مجموعه اعداد طبیعی
$\mathbf{Z}_n$	مجموعه اعداد صحیح به هنگ n
$[a,b]$	مجموعه اعداد حقیقی x با خاصیت $a \leq x \leq b$
$(a,b]$	مجموعه اعداد حقیقی x با خاصیت $a < x \leq b$
(a,b)	مجموعه اعداد حقیقی x با خاصیت $a < x < b$
S^n	مجموعه n تایی‌های مرتب از یک مجموعه S
$A\backslash B$	مجموعه اعضای A که در B قرار ندارند
$\emptyset$	مجموعه خالی
$\|S\|$	اندازه مجموعه S
$n!$	n فاکتوریل
$\binom{n}{r}$	ضریب دوجمله‌ای $(= n!/r!(n-r)!)$
$\approx$	تقریباً برابر است با
$\equiv$	هم‌نهشت است با
$\log a$	لگاریتم a (در یک پایه غیرمشخص)
$\log_r a$	لگاریتم a در پایه r
$\lg a$	$\log_2 a$
$\ln a$	$\log_e a$
∞	بینهایت
$\longrightarrow$	تمایل دارد به، نزدیک می‌شود
$\longmapsto$	نگاشته می‌شود به
$f'(x)$	مشتق تابع $f(x)$
$\wedge, \vee$	رابطه‌های منطقی 'و' و 'یا'
$\cap, \cup$	فصل مشترک و اجتماع
$\sum$	مجموع

$\prod$	ضرب
(a_{ij})	ماتریس با درایه a_{ij} در موقعیت (i,j)
$\det(A)$	دترمینان یک ماتریس A
$\mathrm{tr}(A)$	اثر یک ماتریس A
A^T	ترانهاده یک ماتریس A
I_n	ماتریس $n \times n$ همانی
$\mathbf{a}.\mathbf{b}$	ضرب اسکالر یا نقطه‌ای بردارهای **b** و **a**
$Pr(X_n = s_i)$	احتمال این‌که یک بردار X_n مقدار s_i را اختیار کند، این نماد به‌صورت $Pr(s_i)$ نیز نوشته می‌شود
$Pr(b\|a)$	احتمال b به شرط a
min	می‌نیمم
max	ماکزیمم
$\lfloor x \rfloor$	بخش صحیح x، بزرگ‌ترین عدد صحیح $i \geq x$
$\lceil x \rceil$	تابع سقف، کوچک‌ترین عدد صحیح $i \leq x$

سمبل‌های زیر به‌هنگام اولین استفاده تعریف شده و پس از آن بدون هیچ توضیحی به‌کار می‌روند.

$\mathcal{S}$	یک منبع
S	یک الفبای منبع
s_i	یک سمبل منبع
p_i	احتمال s_i
T	یک الفبای کد
t_j	یک سمبل کد
r	پایه T
w_i	کدکلمه مربوط به s_i
$\mathbf{s}$	یک دنباله از سمبل‌های منبع
$\mathbf{s}$	یک دنباله از سمبل‌های کد
$\|w\|$	طول کلمه w
T^*	مجموعه $\cup_{n\geq 0} T^n$ متشکل از تمامی کلمات روی T
ε	کلمه خالی
T^+	مجموعه تمامی کلمات غیرخالی روی T
$\mathcal{C}$	یک کد منبع
$\mathcal{C}^*$	$\{w_{i_1} w_{i_2} \dots w_{i_n} \in T^* \| w_j \in \mathcal{C}, n \geq 0\}$
l_i	طول کلمه w_i
$L(\mathcal{C})$	میانگین طول کلمه $\sum p_i l_i$ از $\mathcal{C}$
$u.d.$	یکتا دکدپذیر
$\mathcal{C}_n$	مجموعه کلمات $w \in T^+$ که به‌فرم $uw = v$ بوده که $u \in \mathcal{C}$ و $v \in \mathcal{C}_{n-1}$، یا $v \in \mathcal{C}_{n-1}$ و $u \in \mathcal{C}_{n-1}$
$\mathcal{C}_\infty$	$\cup_{n=1}^{\infty} \mathcal{C}_n$
$T^{\leq l}$	$T^0 \cup T^1 \cup T^2 \cup \dots \cup T^l$
$L_{\min}(\mathcal{S})$	بزرگ‌ترین کران پایین برای میانگین طول کلمات کدهای یکتا دکدپذیر r تایی روی $\mathcal{S}$
$\mathcal{S}'$	منبع کاهش‌یافته حاصل از $\mathcal{S}$

$\mathcal{S}^{(i)}$	i امین کاهش $\mathcal{S}$
$\mathcal{C}^{(i)}$	یک کد هافمن برای $\mathcal{S}^{(i)}$
$\sigma(\mathcal{C})$	مجموع $\sum_i l_i$ از طول کلمات $\mathcal{C}$
$\mathcal{S}^n$	n امین بسط منبع $\mathcal{S}$
$\mathcal{C}^n$	یک کد هافمن برای $\mathcal{S}^n$
L_n	$L(\mathcal{C}^n)$
$I(s_i)$	اطلاعات حمل‌شده به‌وسیله s_i ، $-\log p_i$
$I_r(s_i)$	$-\log_r p_i$
$H_r(\mathcal{S})$	آنتروپی r تایی $\mathcal{S}$، $-\sum_i p_i \log_r p_i$
$H(\mathcal{S})$	آنتروپی $\mathcal{S}$، $-\sum_i p_i \log p_i$
$\overline{p}$	$1-p$
$H(p)$	$-p\log p-\overline{p}\log\overline{p}$
$H_r(p)$	$-p\log_r p-\overline{p}\log_r\overline{p}$
η	کارایی $H_r(\mathcal{S})/L(\mathcal{C})$ یک کد r تایی $\mathcal{C}$ برای یک منبع $\mathcal{S}$
$\overline{\eta}$	افزونگی $\eta-1$ از $\mathcal{C}$
$\mathcal{S}\times\mathcal{T}$	ضرب منابع
Γ	یک کانال اطلاعات
$\mathcal{A}$	ورودی Γ
p_i	احتمال سمبل ورودی a_i
q_j	احتمال سمبل خروجی b_j
BSC	کانال دوتایی متقارن
BEC	کانال دوتایی پاک کننده
P_{ij}	احتمال به‌جلو $Pr(b_j\|a_i)$
M	(P_{ij})
$\Gamma+\Gamma'$	مجموع کانال‌ها
$\Gamma\times\Gamma'$	ضرب کانال‌ها
$M\otimes M'$	ضرب کرونکر ماتریس‌ها
Γ^n	بسط n ام کانال Γ
$\Gamma\circ\Gamma'$	ترکیب کانال‌ها
Q_{ij}	احتمال به‌عقب $Pr(a_i\|b_j)$
R_{ij}	احتمال توام $Pr(a_i,b_j)$
$H(\mathcal{A})$	آنتروپی ورودی
$H(\mathcal{B})$	آنتروپی خروجی
$H(\mathcal{A}\|b_j)$	آنتروپی شرطی
$H(\mathcal{A}\|\mathcal{B})$	ایهام
$H(\mathcal{A},\mathcal{B})$	آنتروپی توام
$I(\mathcal{A},\mathcal{B})$	اطلاعات متقابل
C	ظرفیت کانال
$\mathcal{P}$	مجموعه بردارهای توزیع احتمال (p_i) در $\mathbf{R}^r$
Δ	یک قاعده تصمیم $B\to A$
$a_{j\bullet}$	$\Delta(b_j)$

Pr_C	میانگین احتمال کدگشایی درست
Pr_E	میانگین احتمال کدگشایی نادرست
$\mathcal{R}_n$	کد تکراری دوتایی به‌طول n
R	نرخ ارسال
$d(\mathbf{u},\mathbf{v})$	فاصله همینگ بین بردارهای $\mathbf{u}$ و $\mathbf{v}$
F	یک الفبای متناهی که تشکیل یک میدان می‌دهد
p	یک عدد اول
q	یک توانی از عدد اول p
F_q	میدان گالوای از مرتبه q
n	طول یک کد قالبی
$\mathcal{V}=F^n$	یک فضای برداری از بعد n روی F
M	$\lvert\mathcal{C}\rvert$ که $\mathcal{C}$ یک کد است
k	بعد $\dim(\mathcal{C})$ یک کد خطی $\mathcal{C}$
$\mathcal{P}_n$	کد کنترل تساوی به‌طول n روی F_q
$\mathcal{H}_۷$	کد همینگ دوتایی به‌طول ۷
$\mathcal{H}_n$	کد همینگ دوتایی به‌طول n که $n=۲^c-۱$
$\overline{\mathcal{C}}$	کد بسط‌یافته حاصل از $\mathcal{C}$
$\mathcal{C}^\circ$	یک کد پنچر شده حاصل از $\mathcal{C}$
$d=d(\mathcal{C})$	می‌نیمم فاصله $\mathcal{C}$
$wt(\mathbf{v})$	وزن یک بردار $\mathbf{v}$
t	تعداد خطاهای تصحیح شده به‌وسیله یک کد
$\mathbf{e}$	یک بردار خطا
$S_t(\mathbf{u})$	کره به شعاع t و مرکز $\mathbf{u}$
$A_q(n,d)$	بیشترین اندازه ممکن برای یک کد به‌طول n و می‌نیمم فاصله d روی F_q
$\mathcal{C}_۱\oplus\mathcal{C}_۲$	$\{(\mathbf{x},\mathbf{y})\in\mathcal{V}_۱\oplus\mathcal{V}_۲ \lvert \mathbf{x}\in\mathcal{C}_۱,\mathbf{y}\in\mathcal{C}_۲\}$
$\mathcal{C}_۱*\mathcal{C}_۲$	$\{(\mathbf{x},\mathbf{x}+\mathbf{y})\in\mathcal{V}_۱\oplus\mathcal{V}_۲ \lvert \mathbf{x}\in\mathcal{C}_۱,\mathbf{y}\in\mathcal{C}_۲\}$
G	یک ماتریس مولد یک کد $\mathcal{C}$
$\mathbf{e}_i$	i امین بردار مبنای استاندارد
H	یک ماتریس کنترل تساوی برای یک کد $\mathcal{C}$
$\mathcal{C}^\perp$	دوگان یک کد $\mathcal{C}$
c	عدد $n-k$، تعداد ارقام کنترل در یک $[n,k]-$ کد خطی
$\mathcal{G}_{۱۱}$	کد گلی سه‌تایی به‌طول ۱۱
$\mathcal{G}_{۲۳}$	کد گلی دوتایی به‌طول ۲۳
$\mathcal{G}_{۱۲}$	کد گلی سه‌تایی بسط‌یافته $\overline{\mathcal{G}}_{۱۱}$ به‌طول ۱۲
$\mathcal{G}_{۲۴}$	کد گلی دوتایی بسط‌یافته $\overline{\mathcal{G}}_{۲۳}$ به‌طول ۲۴
$\mathcal{P}(S)$	مجموعه توان یک مجموعه S
$S(l,m,n)$	یک سیستم اشتاینر
$PG(c-۱,q)$	هندسه تصویری به بعد $c-۱$ روی F_q
$M_{۲۴}$	گروه ماتیو از درجه ۲۴
$\mathcal{RM}(r,m)$	کد رید—مولر از مرتبه r و به‌طول ۲^m

واژه‌نامه

Alphabet	الفبا ۵۵، ۹۷
Automorphism	اتومورفیسم ۱۴۶، ۱۴۷
Average word-length	متوسط‌طول کلمه ۴، ۴۲
Backward probabilities	احتمال‌های به‌عقب ۵۹
Bayes' formula	فرمول بیز ۶۰
Bernoulli trials	آزمایش‌های برنولی ۱۵۷، ۱۶۰
Binary channel, general	کانال دوتایی، کلی ۷۸
Binary code	کد دوتایی ۳
Binary erasure channel	کانال دوتایی محو کننده ۵۶
Binary Golay code	کد گلی دوتایی ۱۳۷
Binary Huffman code	کد دوتایی هافمن ۲۲، ۲۶، ۲۷، ۲۸
Binary Hamming code	کد دوتایی همینگ ۹۴، ۱۰۲، ۱۳۲
Binary repetition code	کد دوتایی تکراری ۸۵
Binary symmetric channel	کانال دوتایی متقارن ۵۶، ۶۰، ۶۵، ۷۳
Bits	بیت‌ها ۳۷
Block	قالب ۱۳۸، ۱۳۹
Block code	کد قالبی ۵، ۹۹
Block design	طرح قالبی ۱۳۸
Bounded	محدود ۷۶
Capacity	ظرفیت ۷۴
Cascade	آبشار ۵۹
Ceiling function	تابع سقف ۴۵
Channel	کانال ۵۵
Channel capacity	ظرفیت کانال ۷۴
Channel matrix	ماتریس کانال ۵۷
Check digit	رقم(بیت) کنترل ۱۰۰، ۱۰۱، ۱۰۴، ۱۰۹، ۱۱۹، ۱۲۸، ۱۳۰

Group code	کد گروهی
Hadamard code	کد هادامارد ۱۱۷
Hadamard matrix	ماتریس هادامارد ۱۱۴، ۱۱۵
Hamming distance	فاصله همینگ ۸۶
Hamming's sphere-packing bound	کران کره‌چینی همینگ ۱۰۸
Hamming's upper bound	کران بالای همینگ ۱۱۱
Height	ارتفاع ۱۱
Huffman's algorithm	الگوریتم هافمن ۲۲
Huffman code	کد هافمن ۲۲، ۲۶، ۲۷
Ideal observer rule	قاعده مشاهده‌گر ایده‌آل ۸۰
Independent	مستقل ۴۷، ۴۸
Information	اطلاعات ۳۵
Information channel	کانال اطلاعات ۵۵
Information digit	رقم(بیت) اطلاعات ۱۰۰، ۱۰۱، ۱۰۲، ۱۰۴، ۱۱۹، ۱۲۸، ۱۳۰
Information theory	نظریه اطلاعات ۷۹، ۹۷
Input	ورودی ۵۵
Input entropy	آنتروپی ورودی ۶۲
Instantaneous code	کد لحظه‌ای ۸، ۹، ۱۰، ۱۱
Integer part	بخش صحیح
Joint entropy	آنتروپی توام ۶۳
Joint probabilities	احتمال‌های توام ۵۹، ۶۰
Kraft's inequality	نامساوی کرفت ۱۲، ۱۶
Kronecker product	ضرب کرونکر ۵۸
Law of large numbers	قانون اعداد بزرگ ۸۹، ۱۵۷
Leaf	برگ ۱۳
Length	طول ۴
Linear code	کد خطی ۹۹، ۱۲۱
Linear [n, k]-code	[n,k]- کد خطی ۱۰۰
Majority decoding	کدگشایی اکثریت ۸۴
Markov source	منبع مارکف ۵۳
Mathieu group	گروه ماتیو ۱۳۹
Maximum distance separable	جدایی‌پذیر با بیشترین فاصله ۱۱۸
Maximum likelihood rule	قاعده بیشترین درستنمایی ۸۱، ۸۲
McMillan's inequality	نامساوی مک‌میلان ۱۴، ۱۶
Mean value theorem	قضیه مقدار میانی ۶۶
Memoryless	بدون حافظه ۲
Metric space	فضای متری ۸۷

Source code	کد منبع ۳
Source-symbol	سمبل منبع ۵۵
Sphere	کره ۱۰۸
Standard array	آرایش استاندارد ۱۴۰، ۱۴۱
Stationary	پایا ۲
Steiner system	سیستم اشتاینر ۱۳۸، ۱۳۹، ۱۴۰
Stirling's approximation	تقریب استرلینگ ۱۱۱
Strictly convex	اکیداً محدب ۶۵
Sylvester matrix	ماتریس سیلوستر ۱۱۶
Syndrome	مشخصه ۱۳۴، ۱۴۳، ۱۴۴
Syndrome decoding	کدگشایی مشخصه ۱۴۳
Syndrome table	جدول مشخصه ۱۴۴
System entropies	دستگاه آنتروپی ۶۲
Systematic form	فرم منظم ۱۲۷، ۱۲۸، ۱۲۹
t-design	t- طرح ۱۳۸
Ternary code	کد سه‌تایی ۳
Ternary Golay code	کد گلی سه‌تایی ۱۳۷
t-error-correcting	t خطا تصحیح کننده ۱۰۶، ۱۳۲
Transmission rate	نرخ ارسال ۸۵، ۸۶
Tree	درخت ۱۰
Triangle inequality	نامساوی مثلث ۸۷
Uniform	یکنواخت ۷۸
Uniquely decipherable	یکتا رمزگشایی شونده ۴
Uniquely decodable	یکتا دکدپذیر ۴، ۵، ۶، ۷
Uniquely decodable with bounded delay	یکتا دکدپذیر با تاخیر محدود ۸
Vertex	راس ۱۰، ۱۱
Weak law of large numbers	قانون ضعیف اعداد بزرگ ۱۵۸، ۱۶۰
Weight	وزن ۱۰۵
Word	کلمه ۳
Word-length	طول کلمه ۴

Gareth A. Jones, MA, DPhil
Faculty of Mathematical Studies, University of Southampton,
Southampton SO17 1BJ, UK

J. Mary Jones, MA, DPhil
The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

ISSN 1615-2085

ISBN 1-85233-622-6 Springer-Verlag London Berlin Heidelberg

British Library Cataloguing in Publication Data
Jones, Gareth A.
Information and coding theory. - (Springer undergraduate mathematics series)
1. Information theory 2. Coding theory
I. Title II. Jones, J. Mary
003.5'4
ISBN 1852336226

Library of Congress Cataloging-in-Publication Data
Jones, Gareth A.
Information and coding theory / Gareth A. Jones and J. Mary Jones.
p. cm. -- (Springer undergraduate mathematics series)
Includes bibliographical references and index.
ISBN 1-85233-622-6 (alk. paper)
1. Information theory. 2. Coding theory. I. Jones, J. Mary (Josephine Mary), 1946-
II. Title. III. Series.
Q360 .J68 2000
003'.54—dc21 00-030074

Typesetting: Camera ready by the author and Michael Mackey
Printed and bound at the Athenæum Press Ltd., Gateshead, Tyne & Wear
12/3830-543210 Printed on acid-free paper SPIN 10731522

Information and coding Theory

by:

G. A Jones
J. M. Jones

Translated by:

Morteza Esmaeili *Ph.D.*

Assistant Professor
Isfahan University of Technology

2004

INFORMATION AND CODING THEORY

by:

G. A. Jones & J. M. Jones

Translated by:

M. Esmaeili Ph.D.

Assistant Professor of Mathematics

ISBN : 964-6029-97-3