نظریهٔ اطلاع

جان سی. الف. وان دِر لوب

ترجمهٔ : دکتر حسنعلی آذرنوش

۳۰۸

انتشارات دانشگاه فردوسی مشهد، شمارهٔ ۳۰۸

نظریهٔ اطلاع

جان سی. الف. وان دِر لُوب

ترجمهٔ

دکتر حسنعلی آذرنوش

۱۳۸۰

لوبه، یان واندر، Lubbe, Jan C. A. Vander

نظریه اطلاع / جان سی. الف. وان در لوب؛ ترجمه حسنعلی آذرنوش. ـ مشهد: دانشگاه فردوسی مشهد، ۱۳۸۰.

۳۹۸ ص.: مصور، جدول، نمودار. ـ (انتشارات دانشگاه فردوسی مشهد؛ ۳۰۸).

۱۸۰۰۰ ریال (ISBN: 964-5782-35-x)

فهرست‌نویسی بر اساس اطلاعات فیپا.

عنوان اصلی: Information theorie = Information theory.

واژه‌نامه.

کتابنامه: ص. [۳۷۹] - ۳۸۰.

۱. نظریه اطلاعات. الف. آذرنوش، حسنعلی، ۱۳۱۹ - : مترجم. ب. دانشگاه فردوسی (مشهد). ج. عنوان.

۶ ن ۹ ل / Q ۳۶۰ ۰۰۳/۵۴

۱۳۸۰

کتابخانه ملی ایران م ۸۰-۱۶۰۶۵

نظریهٔ اطلاع

نوشتهٔ

لوبه، یان واندر

ترجمهٔ

حسنعلی آذرنوش

وزیری، ۴۰۰ صفحه، ۱۰۰۰ نسخه، چاپ اوّل، پاییز ۱۳۸۰

امور فنّی و چاپ : مؤسسهٔ چاپ و انتشارات دانشگاه فردوسی

بها : ۱۸۰۰۰ ریال

شابک ۹۶۴-۵۷۸۲-۳۵-x (ISBN: 964-5782-35-x)

فهرست

عنوان	صفحه

۳ منبع اطلاع گسسته باحافظه

۴ کانال ارتباطی گسسته

۵ منبع اطلاع پیوسته

پیش گفتار

در تمام سطوح جامعه سیستمهایی مرسوم شده‌اند که با انتقال، ذخیره‌سازی و پردازش اطلاعات سروکار دارند. ما در جامعه‌ای زندگی می‌کنیم که معمولاً جامعهٔ اطلاعاتی نامیده می‌شود. اطلاع در جامعهٔ ما شکل کلیدی به خود گرفته است؛ بنابراین حیرت انگیز نیست که تمام بخشهای مختلف در دانستن این‌که اطلاعات در حقیقت چیست و در نتیجه در کسب دانش بیشتر در جهت کاربرد اطلاعات، به‌گونه‌ای که تا حدّ امکان مؤثر باشد، از خود تمایل نشان می‌دهند.

نظریهٔ اطلاع با توجّه به مفهوم اطلاع به شیوه‌ای کمّی توصیف می‌شود. به منظور معرّفی اندازه‌ای برای اطلاع، درصدد پاسخ به پرسشهایی چون: چگونه اطلاعات را وقتی در حدّ امکان فشرده شده‌اند منتقل و ذخیره کنیم؟ حداکثر مقدار اطلاع که می‌توان از طریق یک کانال ارسال کرد چقدر است؟ چگونه می‌توان محافظت را به بهترین نحو ترتیب داد؟ و یا دیگر پرسشها، خواهیم بود. سؤالات کلیدی، ما را در درک بهتر محدودیتهای سیستم یاری می‌دهند.

این کتاب در نظر دارد تعدادی از مفاهیم اساسی نظریهٔ اطلاع را معرّفی کند و با نشان دادن اهمیت آنها در کاربرد موجود به تبیین آنها بپردازد. مسائلی که مطرح خواهند شد از میان دیگر مطالب، «اندازهٔ اطلاع» شانون[1]، منابع گسسته و پیوستهٔ اطلاع و کانالهای اطلاع با حافظه و بدون حافظه، رمزگشایی منبع و کانال، نظریهٔ نرخ دگرشکلی، کدهای تصحیح خطا و شیوهٔ نظری اطلاع در علم رمزشناسی می‌باشند. توجّه خاصی به نظریهٔ چندپایانه‌ای یا شبکه‌ای اطلاعات شده است. مبحثی با پرسشهای بی‌پاسخ فراوان امّا بسیار مهم چرا که بیشتر اطلاعات به وسیلهٔ شبکه‌ها ارسال می‌شوند.

تمام فصلها به پرسشها و راه حلهای کار شده ختم می‌شوند. این امر، کتاب را برای مطالعهٔ فردی مناسب می‌سازد.

متن این کتاب عمدتاً بر پایهٔ سخنرانیهای اخیر نویسنده برای دانشجویان مهندسی الکترونیک، ریاضیات تکنیکی و انفورماتیک، فیزیک کاربردی و مهندسی مکانیک در دانشگاه تکنولوژی دُلف و همچنین متن سخنرانیهای سابق استادان یسبراند باکسما[2]، دیک

1. Shannon. 2. Ysbrand Boxma.

بوکی[1] و جان بیموند[2] می‌باشد. پرسشها از امتحانهای اخیر گرفته شده‌اند.

نویسنده مایل است مراتب قدردانی خود را از همکاران فوق‌الذکر بـه عـلاوهٔ دیگـر همکاران که به هر نحو به نوشتن این کتاب یاری رساندند، ابراز دارد. مخصوصاً مایلم کـه از یسبراند باکسما که با سخنرانی خود در زمینهٔ نظریهٔ اطلاع در دانشگاه تکنولوژی دلـف در زمانی که من دانشجو بودم، مرا با نظریهٔ اطلاع آشـنا سـاخت سـپاس‌گزاری کنـم. بـا راهنماییهای الهام بخش او من درجهٔ کارشناسی ارشد در مهندسی الکترونیک و دکتری در علوم تکنولوژی دریافت کردم. در نوشتن این کتاب متنهای سخنرانیهای سابق او بسیار بـه من کمک کرد. تأثیر او یک عامل تعیین کننده در حرفهٔ اخیر من بوده است.

جان سی. الف. وان دِر لُوب[3]

دُلف، دسامبر ۱۹۹۶

1. Dick Boekee. 2. John Biemond 3. Jan C.A. van der Lubbe.

۱

اطلاع گسسته

۱.۱ مبدأ نظریهٔ اطلاع

نظریهٔ اطلاع علمی است که با مفهوم اندازه و کـاربرد «اطـلاع» سـروکار دارد. بـه مفهوم وسیع آن بین سنتهای امریکایی و انگلیسی در نظریهٔ اطلاع می‌توان وجه تمایزی در نظر گرفت.

به طورکلّی سه نوع اطلاع وجود دارد:

- **اطلاع ترکیبی،** به علاماتی که پیامها با آنها ساخته می‌شـوند و رابطـهٔ بیـن آنهـا وابسته است.
- **اطلاع معانی،** وابسته به معانی پیامها و جنبه‌های معرّفی آن می‌باشد.
- **اطلاع عملی،** وابسته به کارگیری و اثر پیامهاست.

طبیعت آنها چنین است، اطلاع ترکیبی اصولاً شکل اطـلاع را در نظـر می‌گـیرد؛ در حالی که اطلاع معانی و عملی وابسته به محتوای اطلاع می‌باشند.

جملات زیر را در نظر بگیرید:

(الف) شخصی با تاکسی به راه‌آهن آمد.

(ب) تاکسی شخصی را به راه‌آهن آورد.

(پ) ترافیک سنگینی در بزرگراه $A3$، بین نورنبرگ و مونیخ در آلمان وجود دارد.

(ت) ترافیک سنگینی در بزرگراه $A3$ در آلمان وجود دارد.

جملات (الف) و (ب) به طور ترکیبی متفاوتند. با وجود این به صورت معانی وعلمی آنها یکی هستند. آنها دارای یک معنی هستند و هر دو دارای اطلاع برابرند.

جملات (پ) و (ت) نه تنها از نظر ترکیبی متفاوتند، بلکه همچنین نسبت به معانی یکی نیستند. جملهٔ (پ) اطلاع دقیقتری را از جملهٔ (ت) می‌دهد.

اصولاً جنبهٔ عملی اطلاع وابسته به متن می‌باشد. برای مثال محتوای اطلاع در جملات (پ) و (ت) مربوط به فردی است که در آلمان می‌باشد و نه کسی که در امریکاست.

جنبه‌های مفهومی و عملی اطلاع مطالعه شده به سنت انگلیسی در نظریهٔ اطلاع این چنین است. سنت انگلیسی وابستگی نزدیکی به فلسفه، روان‌شناسی و بیولوژی دارد. سنت انگلیسی اصولاً متأثر از دانشمندانی نظیر مک کی[1]، کارناپ[2]، بار- هیلل[3]، آکوف[4] و هینتیکا[5] می‌باشد.

سنت امریکایی با جنبه‌های ترکیبی اطلاع سروکار دارد. در این روش، تجرّد کاملی از معانی جنبه‌های اطلاع وجود دارد. سؤالهای اساسی عبارتند از: اندازهٔ اطلاع ترکیبی، حدود اساسی بر مقدار اطلاعی که می‌توان ارسال کرد، حدود اساسی بر فشردگی اطلاعی که می‌توان به دست آورد و چگونه سیستمهای پردازش اطلاع را برای رسیدن به این حدود بسازیم. چیزی که باقی می‌ماند یک رهیافت نسبتاً تکنیکی برای رسیدن به اطلاع است.

گاهی اوقات سنت امریکایی در نظریهٔ اطلاع به عنوان نظریهٔ ارتباطات، نظریهٔ اطلاع ریاضی، یا به طور خلاصه نظریهٔ اطلاع در نظر گرفته می‌شود. دانشمندان مشهور سنت امریکایی در بین دیگران عبارتند از: شانون، رنی[6]، گالاگر[7]، و سیساب[8].

با وجود این، کلود ای. شانون که او مَقالـه‌اش را تحت عنوان «نظریهٔ ریاضی ارتباطات» در ۱۹۴۸ چاپ کرد عموماً به عنوان مؤسّس سنت امریکایی در نظریهٔ اطلاع شناخته می‌شود. با این همه، تعدادی پیشرو نسبت به شانون وجود دارند که کوشیده‌اند کارایی استفاده از سیستم ارتباطات را فرمول‌بندی کنند.

اچ. نیکویست[9] در سال ۱۹۲۴ مقاله‌ای چاپ کرد که در آن چگونگی ارسال پیامها (یا نوشته‌ها، با استفاده از خود کلمات) را توسط یک کانال تلگراف با ماکسیمم سرعت ممکن ولی بدون دگرشکلی فراهم نمود. با وجود این جمله اطلاع هنوز توسط او این چنین استفاده نشده است.

---

1. MacKay
2. Carnap
3. Bar-Hillel
4. Ackoff
5. Hintikka
6. Renyi
7. Gallager
8. Csiszar
9. H.Nyquist

ار. وی. ال. هارتلی[1] (۱۹۲۸) اوّلین فردی است که کوشید ***اندازهٔ اطلاع*** را تعریف کند. او در این باره به طریق زیر اقدام کرد

فرض کرد که هر نماد یک پیام را بتوان به s طریق انتخاب کرد؛ اکنون با در نظر گرفتن پیامهای l نمادی می‌توان s^l پیام متمایز تشخیص داد. اینک هارتلی مقدار اطلاع را به صورت لگاریتم تعداد پیامهای قابل تشخیص تعریف می‌کنند. بنابراین در حالتی که پیامها با طول l باشند داریم

$$H_H(s^l) = \log\{s^l\} = l\log\{s\} \qquad (۱.۱)$$

برای پیامهای به طول یک داریم

$$H_H(s^1) = \log\{s\}$$

در نتیجه می‌توان نوشت

$$H_H(s^l) = l\,H_H(s^1)$$

این نتیجه با درک ذهنی این‌که اطلاع هر پیام به طول l، l برابر اطلاع پیامی به طول یک است سازگار می‌باشد. این مطلب حضور موجه لگاریتم را در تعریف هارتلی نیز بیان می‌کند.

به سادگی می‌توان نشان داد که تنها تابعی که در معادلهٔ

$$f\{s^l\} = l\,f\{s\}$$

صدق می‌کند به صورت زیر است

$$f\{s\} = \log\{s\} \qquad (۲.۱)$$

که اندازهٔ هارتلی برای مقدار اطلاع حاصل می‌گردد. توجّه کنید که لگاریتم با افزایش تعداد نمادهای s افزایش مقدار اطلاع را نیز تضمین می‌کند که با درک مستقیم تطابق دارد.

انتخاب مبنای لگاریتم دلخواه است و بیشتر یک موضوع عادی‌سازی می‌باشد. اگر از لگاریتم طبیعی استفاده شود، واحد اطلاع را **نَت** (*nat*) (واحد طبیعی) می‌نامند. معمولاً عدد ۲ را به عنوان مبنا انتخاب می‌کنند. در این صورت واحد اطلاع برحسب **بیت** (حاصل از دستگاه دودویی، یعنی دستگاه دو مقداری) بیان می‌شود. در حالت انتخاب دو امکانی وقتی یکی از دو امکان رخ دهد مقدار اطلاع در این صورت برابر یک بیت است. به سادگی دیده می‌شود که رابطهٔ بین بیت و نَت به صورت زیر است

1. R.V.L.Hartley

۱٫۴۴ بیت = ۱ نَت

در رهیافت هارتلی همان‌طور که در بالا ذکر شد هیچ فرضی در این مورد که امکان دارد s نماد با شانسهای نابرابر رخ دهند یا این که ممکن است بستگی بین نمادهای متوالی وجود داشته باشد در نظر گرفته نشده است.

دستاورد بزرگ شانون این است که او نظریه‌های نیکویست و هارتلی را توسعه داد و نظریهٔ اطلاع امروزی را با مرتبط ساختن اطلاع با عدم‌حتمیت با بهره‌وری از مفهوم شانس یا احتمال پایه‌گذاری کرد. راجع به اندازه هارتلی، شانون پیشنهاد کرد به فرض این که همهٔ نمادها با احتمال برابر رخ دهند آن را می‌توان واقعاً به عنوان اندازه اطلاع تفسیر کرد. در حالت کلّی شانون اندازهٔ اطلاع را بر مبنای مفهوم احتمال معرّفی کرد که اندازهٔ هارتلی را به عنوان حالت خاصی شامل می‌شود. قبل از معرّفی تعریف اطلاع شانون ابتدا به نظریهٔ احتمال توجّه خواهد شد که در ضمن آن مفاهیم مفیدی معرّفی خواهیم کرد.

۲.۱ مفهوم احتمال

نظریهٔ احتمال زمینه‌ای در علوم است که با مفهوم احتمال سروکار دارد. نقطهٔ شروع نظریهٔ احتمال انجام آزمایشهایی است که منجر به برآمدهایی می‌شود. همچنین می‌توان برحسب منبع اطلاعاتی که نمادها را تولید می‌کند در نظر گرفت. در این صورت هر رخ‌داد یک نماد را می‌توان به عنوان یک پیشامد در نظر گرفت. فرض بر این است که قادر به تشخیص برآمدها یا پیشامدهای ممکنی که می‌تواند رخ دهد می‌باشیم، مجموعه همهٔ برآمدها یا پیشامدهای ممکن را ***فضای نمونه*** می‌نامند. اکنون ممکن است از احتمال این‌که یک آزمایش دارای یک برآمد خاص است یا احتمال این‌که یک منبع اطلاع یک نماد یا پیام خاصی را تولید خواهد کرد صحبت کرد. به هر پیشامد یا برآمد عددی بین ۰ و ۱ نسبت می‌دهیم، که احتمال رخ‌داد این برآمد یا پیشامد را مشخّص می‌کند. برای سادگی، فرض شده است که فضای نمونه دارای تعداد معیّنی برآمد است.

آزمایش احتمالی X با برآمدها یا پیشامدهای ممکن x_i، $x_i \in X$ و فضای احتمال X به صورت زیر را در نظر بگیرید

$$X = \{x_1, \ldots, x_j, \ldots, x_n\}. \qquad (۳.۱)$$

اگر تصوّر کنیم که یک تاس پرتاب شده است، در این صورت x_1 را می‌توان به

عنوان پیشامدی که "۱" آمده است، x_2 پیشامدی که "۲" آمده است و الی‌آخر تفسیر نمود. در حالت پرتاب تاس واضح است که $n = 6$ می‌باشد.

هر پیشامدی احتمال رخ‌داد معینی خواهد داشت. احتمال مربوط به x_i را با $p(x_i)$ و یا به طور ساده با p_i نشان می‌دهیم. مجموعهٔ احتمالها دربارهٔ X را به صورت زیر نشان می‌دهیم

$$P = \{p_1, \ldots, p_i, \ldots, p_n\}, \qquad (4.1)$$

و آن را ***توزیع احتمال*** می‌نامیم. توزیع احتمال در دو شرط اساسی زیر صدق می‌کند:

(الف) برای همهٔ iها، $p_i \geq 0$.

(ب) $\sum_{i=1}^{n} p_i = 1$.

یعنی هیچ احتمالی نمی‌تواند مقدار منفی اختیار کند و مجموع همهٔ احتمالها برابر یک است.

گاهی می‌توان دو نوع برآمد در یک آزمایش تشخیص داد، به طوری که ترکیبی از دو آزمایش فرعی یا پیشامد فرعی داریم. برای مثال، وقتی ICها را می‌آزماییم می‌توان به این‌که تا چه حدّی از شرایط معیّن (برای مثال، خوب، متوسط، بد) حاصل شده است و همچنین به تعداد گونهٔ ICها توجّه کرد. در این صورت به واقع با دو فضای نمونه مانند X و Y سروکار داریم که در آن فضای نمونهٔ Y وابسته به آزمایش Y را به طور کلّی به صورت زیر تعریف می‌کنیم

$$Y = \{y_1, \ldots, y_j, \ldots, y_m\}, \qquad (5.1)$$

با توزیع احتمال

$$Q = \{q_1, \ldots, q_j, \ldots, q_m\}, \qquad (6.1)$$

که در آن $q(y_j) = q_j$ احتمال پیشامد y_j می‌باشد. اکنون می‌توان (X,Y) را به صورت آزمایش احتمالی با زوج برآمدهای (x_i, y_j) با $x_i \in X$ و $y_j \in Y$ در نظر گرفت، احتمال $r(x_i, y_j)$ که با r_{ij} یا $p(x_i, y_j)$ نیز نشان داده می‌شود برابر احتمال آزمایش (X,Y) است که برآمد (x_i, y_j) را خواهد داد و آن را ***توزیع توأم*** می‌نامیم. اگر توزیع توأم معلوم باشد می‌توان احتمالهای p_i و q_j را به دست آورد و آنها را ***احتمالهای حاشیه‌ای*** می‌نامیم. می‌توان ثابت کرد که برای تمام iها داریم

$$p_i = \sum_{j=1}^{m} r_{ij}, \qquad (7.1)$$

و برای همهٔ jها داریم

$$q_j = \sum_{i=1}^{n} r_{ij} \,. \qquad (۸.۱)$$

چون مجموع تمام احتمالهای p_i باید برابر یک باشد (و به طور مشابه مجموع احتمالهای q_j)، نتیجه می‌شود که مجموع احتمالهای توأم نیز باید برابر یک باشد:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} = 1.$$

علاوه بر احتمال توأم و احتمال حاشیه‌ای وابسته به آن، نوع سومی یعنی ***احتمال شرطی*** وجود دارد. این نوع احتمال وقتی آزمایش احتمالی Y به شرط X است حاصل می‌گردد. یعنی احتمالهای برآمدهای X تحت تأثیر برآمدهای Y می‌باشند. در این صورت به احتمال پیشامدی مانند x_i به شرط پیشامد دیگری مانند y_j که قبلاً رخ داده است، علاقه‌مندیم.

واژه‌ای را در یک قطعه انگلیسی در نظر بگیرید. برای مثال، اگر قبلاً دنبالهٔ "*informatio*" را داشته باشیم ممکن است از خود بپرسیم احتمال این‌که حرف "*n*" ظاهر شود چقدر است. حضور حروف در کلمات اغلب بستگی به حروفی دارد که قبلاً ظاهر شده‌اند. برای مثال، خیلی غیرمحتمل است که حرف "*q*" با حرف "*t*" دنبال شود ولی خیلی بیشتر محتمل است که با حرف "*u*" دنبال شود.

احتمال شرطی x_i به شرط y_j به صورت زیر تعریف می‌شود:

$$p(x_i \mid y_j) = \frac{r(x_i, y_j)}{q(y_j)}, \quad q(y_j) > 0 \text{ به شرط}$$

یا به طور اختصار به صورت

$$p_{ij} = \frac{r_{ij}}{q_j}, \quad q_j > 0 \text{ به شرط} \qquad (۹.۱)$$

احتمال y_j به شرط x_i به طور مشابه عبارت است از

$$q(y_j \mid x_i) = \frac{r(x_i, y_j)}{p(x_i)}, \quad p(x_i) > 0 \text{ به شرط}$$

یا به طور ساده

$$q_{ji} = \frac{r_{ij}}{p_i}, \quad p_i > 0 \text{ به شرط} \qquad (۱۰.۱)$$

از تعاریف داده شده نتیجه می‌شود که احتمال توأم را می‌توان به صورت حاصل‌ضرب

احتمالهای شرطی و حاشیه‌ای نوشت

$$r(x_i,y_j)=q(y_j)p(x_i\,|\,y_j)=p(x_i)q(y_j\,|\,x_i). \qquad (۱۱.۱)$$

تعریف احتمال شرطی را به طور ساده می‌توان برای بیش از دو پیشامد گسترش داد. برای مثال x_i، y_j و z_k را در نظر بگیرید

$$\begin{aligned} p(x_i,y_j,z_k) &= r(y_j,z_k)p(x_i\,|\,y_j,z_k) \\ &= p(z_k)p(y_j\,|\,z_k)p(x_i\,|\,y_j,z_k), \end{aligned}$$

از این رو

$$p(x_i\,|\,y_j,z_k)=\frac{p(x_i,y_j,z_k)}{r(y_j,z_k)}.$$

با مراجعه به احتمال شرطی، با جمع‌بندی روی زیرنویس i به شرط y_j داریم

$$\sum_{i=۱}^{n} p(x_i\,|\,y_j)=۱. \qquad (۱۲.۱)$$

یعنی هرگاه یک پیشامد y_j رخ داده باشد یکی از پیشامدها در X نیز بایستی رخ دهد. از این رو با جمع‌بندی ۱ به دست خواهد آمد. توجّه کنید که عکس آن درست نیست. به طور کلّی درست است که

$$\sum_{j=۱}^{m} p(x_i\,|\,y_j)\neq ۱. \qquad (۱۳.۱)$$

قضیهٔ مفیدی که در ذیل، مورد استفاده خواهد بود ***قضیهٔ بیز*** است؛ اغلب با معلوم بودن احتمال شرطی $q(y_j\,|\,x_i)$ می‌خواهیم احتمال شرطی $p(x_i\,|\,y_j)$ را تعیین کنیم. برای انجام این کار می‌توان از روابط زیر استفاده نمود

$$r(x_i,y_j)=p(x_i)q(y_j\,|\,x_i)=q(y_j)p(x_i\,|\,y_j).$$

بنابراین اگر $q(y_j)>۰$ داریم

$$p(x_i\,|\,y_j)=\frac{p(x_i)q(y_j\,|\,x_i)}{q(y_j)},$$

یا همچنین داریم

$$p(x_i\,|\,y_j)=\frac{p(x_i)q(y_j\,|\,x_i)}{\sum_{i=۱}^{n} p(x_i)q(y_j\,|\,x_i)}. \qquad (۱۴.۱)$$

از این رو می‌توانیم با کمک $q(y_j|x_i)$ها مقدار $p(x_i|y_j)$ را حساب کنیم.

در پایان تفسیری در مورد مفهوم استقلال داریم. وضعیت به صورت زیر رخ می‌دهد

$$p(x_i|y_j)=p(x_i).$$

یعنی، رخ دادن y_j هیچ‌گونه تأثیر بر رخ‌داد x_i ندارد. همچنین نتیجه می‌شود که:

$$r(x_i,y_j)=p(x_i)q(y_j)$$

و

$$q(y_j|x_i)=q(y_j).$$

در این حالت می‌توان گفت که پیشامدها از هم مستقلند. عکس آن نیز درست است. از $r(x_i,y_j)=p(x_i)q(y_j)$ نتیجه می‌شود که $q(y_j|x_i)=q(y_j)$ و $p(x_i|y_j)=p(x_i)$. دو آزمایش X و Y را به طور آماری مستقل گوییم اگر و تنها اگر برای تمام i و jها داشته باشیم

$$r(x_i,y_j)=p(x_i)q(y_j). \tag{۱۵.۱}$$

یک آزمایش X را با آزمایش دیگر Y کاملاً وابسته گوییم، اگر برای تمام jها یک i منحصر به فردی مانند k وجود داشته باشد به طوری که

$$p(x_k|y_j)=1, \tag{۱۶.۱}$$

یا

$$p(x_k,y_j)=p(y_j). \tag{۱۷.۱}$$

۳.۱ اندازهٔ اطلاع شانون

همان‌طور که در بخش (۱.۱) دیدیم، تعریف اطلاع هارتلی رخ‌داد نمادها یا پیشامدها را با احتمالهای متفاوت در نظر نمی‌گیرد. شانون اوّلین کسی است که اطلاع را با مفهوم احتمال پیوند داد.

در حقیقت این ارتباط غیرمنطقی نیست. اگر فضای نمونه‌ای را که در آن همهٔ پیشامدها دارای احتمال رخ‌داد برابرند در نظر بگیریم، عدم‌حتمیت زیادی دربارهٔ این که کدام پیشامد رخ خواهد داد وجود دارد. یعنی وقتی یکی از این پیشامدها رخ می‌دهد اطلاع بیشتری فراهم می‌کند از حالتهایی که در آن فضای نمونه به طریقی ساخته شده که یک

پیشامد با احتمال بالایی رخ می‌دهد. اطلاع از طریق عدم‌حتمیت با مفهوم شانس پیوند می‌خورد.

قبل از این که بررسی کنیم که تا چه حدّی *اندازهٔ شانون* ویژگیهایی را که عموماً از اندازهٔ اطلاع انتظار داریم داراست، ابتدا تعریف او را بیان می‌کنیم

تعریف ۱.۱

فرض کنید X یک آزمایش احتمالی با فضای نمونهٔ X و توزیع احتمال P باشد، که در آن $p(x_i)$ یا p_i احتمال برآمد $x_i \in X$ می‌باشد. در این صورت متوسط مقدار اطلاع به صورت زیر داده می‌شود

$$H(\mathrm{X}) = -\sum_{i=1}^{n} p(x_i)\log p(x_i) = -\sum_{i=1}^{n} p_i \log p_i \,. \qquad (۱۸.۱)$$

نمادهای دیگر برای اندازهٔ اطلاع شانون عبارتند از: $H(X)$، $H(P)$ و $H(p_1,\ldots,p_n)$.
تمام این نمادها در این کتاب به طور قابل تعویضی به کار برده خواهد شد، زیرا این اندازه برای مقدار اطلاع با انتخاب از n امکان ناشی می‌شود، گاهی به آن اندازهٔ مقدار *اطلاع انتخابی* نیز گفته می‌شود.

چون معمولاً عدد ۲ به عنوان مبنا انتخاب می‌شود بنابراین واحد اطلاع بیت می‌باشد. این مطلب را در آینده دیگر تکرار نخواهیم کرد. در حالت دو برآمد با احتمالهای $p_1 = p$ و $p_2 = 1 - p$ داریم

$$H(P) = -p\log p - (1-p)\log(1-p) \qquad (۱۹.۱)$$

شکل (۱.۱) نشان می‌دهد که چگونه $H(P)$ به عنوان تابعی از p عمل می‌کند. می‌توان نتیجه گرفت که اگر یک برآمد حتمی باشد، یعنی با احتمال یک رخ دهد، اندازهٔ اطلاع صفر به دست می‌آید. این مطلب با درک مستقیم ذهنی که پیشامدهای حتمی هیچ اطلاعی را فراهم نمی‌کنند موافق است. این مطلب برای $p = 0$ نیز درست است؛ در این حالت برآمد دیگر دارای احتمال یک است.

وقتی $p = 0.5$، $H(P)$ به ماکسیمم مقدار خود می‌رسد، که برابر با یک بیت است. برای $p = 0.5$ هر دو برآمد دارای احتمال یکسانند، و فرد در مورد برآمد کاملاً نامطمئن می‌باشد. در این حالت رخ‌داد یکی از پیشامدها ماکسیمم مقدار اطلاع را فراهم می‌کند.

در ضمن، توجّه کنید که بنا به تعریف داریم: $0 \times \log(0) = 0$.

با بازگشت به حالت کلّی می‌توان فرض کرد که اندازهٔ اطلاع در چهار شرط بدیهی

زیر صدق می‌کند:

۱- تابع $H(P)$ برحسب p **پیوسته** است.

۲- تابع $H(P)$ **متقارن** است. یعنی، ترتیب احتمالهای p_1، ...، p_n بر مقدار $H(P)$ تأثیری ندارد.

۳- تابع $H(P)$ **جمع‌پذیر** است. اگر X و Y دو فضای نمونه باشند به قسمی که برآمدهای X مستقل از برآمدهای Y باشند، آن گاه برای اطلاع وابسته به پیشامدهای توأم (x_i, y_j) داریم

$$H(p_1q_1, \ldots, p_1q_m, \ldots, p_nq_1, \ldots, p_nq_m) = H(p_1, \ldots, p_n) + H(q_1, \ldots, q_m). \qquad (۱.۲۰)$$

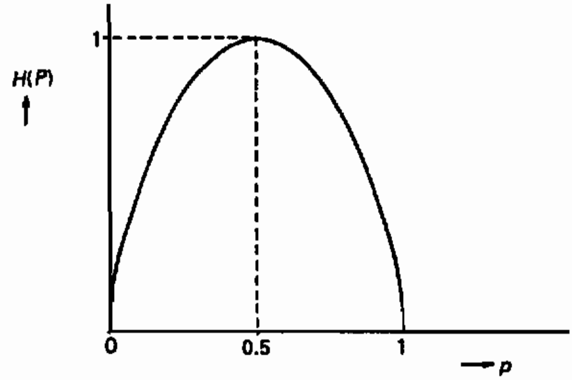
۴- اگر تمام احتمالها مساوی باشند $H(P)$ ماکسیمم می‌شود. این مطلب با حالتی متناظر است که در آن عدم‌حتمیت ماکسیمم وجود دارد. اگر برآمد دارای احتمال یک باشد مقدار $H(P)$ می‌نیمم می‌شود.

تفسیر کوتاهی از بعضی از شرایط بالا به صورت زیر است.

تفسیر ۲- تقارن اندازه اطلاعات شانون بدین معنی است که تعویض ترتیب احتمالها مقدار اطلاع را تغییر نمی‌دهد. یک نتیجهٔ این مطلب این است که مقدار اطلاع فضاهای نمونهٔ متفاوت با توزیع احتمالی که از جایگشت توزیع احتمال مشترک به دست می‌آید یکسانند.

مثال ۱.۱

آزمایشهای X و Y با فضاهای نمونهٔ زیر را در نظر بگیرید:



شکل ۱.۱- $H(P) = H(p, 1-p)$ به عنوان تابعی از p

$X =$ { فردا باران می‌بارد ، فردا باران نمی‌بارد }

که در آن $P = \{0.8, 0.2\}$ و

$Y =$ { حامد کمتر از ۳۰ سال دارد ، حامد حداقل ۳۰ سال دارد }

که در آن $Q = \{0.2, 0.8\}$.

مقدار اطلاع در رابطه با X عبارت است از

$$H(\mathrm{X}) = -0.8\log 0.8 - 0.2\log 0.2 = 0.72 \text{ بیت}$$

و در رابطه با Y برابر است با

$$H(\mathrm{Y}) = -0.2\log 0.2 - 0.8\log 0.8 = 0.72 \text{ بیت}$$

و از این رو داریم

$$H(\mathrm{X}) = H(\mathrm{Y}). \qquad \blacktriangle$$

از این مثال می‌توان نتیجه گرفت که اندازهٔ اطلاع شانون به محتوای اطلاع بستگی ندارد. احتمال رخ دادن پیشامدها مهم هستند نه خود پیشامدها.

تفسیر ۳- اندازهٔ اطلاع شانون در خاصیت (۱. ۲۰) صدق می‌کند که مستقیماً با نوشتن آن برحسب احتمالها نتیجه می‌گردد. ویژگی جمع‌پذیری با مثال زیر با بهترین وجهی توضیح داده می‌شود. دو تاس در نظر بگیرید. چون برآمدهای دو تاس مستقل از یکدیگر می‌باشند، فرقی نمی‌کند که دو تاس هم‌زمان یا یکی بعد از دیگری پرتاب شود. اطلاع وابسته به تاسها وقتی با هم پرتاب می‌شوند همانند اطلاع متوالی است که با پرتاب تاسها یکی پس از دیگری به دست می‌آید.

اگر $H(\mathrm{X})$ مقدار اطلاع مربوط به پرتاب یک تاس و $H(\mathrm{Y})$ مقدار اطلاع مربوط به پرتاب تاس دیگر باشد (توجّه کنید در این حالت $H(\mathrm{X}) = H(\mathrm{Y})$)، در حالی که $H(\mathrm{X},\mathrm{Y})$ اطلاع مربوط به پرتاب دو تاس هم‌زمان باشد آن‌گاه بایستی نتیجه شود که

$$H(\mathrm{X},\mathrm{Y}) = H(\mathrm{X}) + H(\mathrm{Y}) \qquad (۲۱.۱)$$

این دقیقاً همان چیزی است که ویژگی جمع‌پذیری بیان می‌کند.

تفسیر ۴- واضح است که مقدار اطلاع در حالت احتمالهای برابر ماکسیمم خواهد شد، در این صورت از نقطه نظر این حقیقت که عدم‌حتمیت بیشترین است رخ‌داد یکی از

پیشامدها ماکسیمم اطلاع را نتیجه خواهد داد.

در قضیهٔ زیر علاوه بر مقدار ماکسیمم اطلاع مقدار می‌نیمم اطلاع نیز تعیین می‌شود.

قضیهٔ ۱.۱

فرض کنید $X=(x_1,\ldots,x_n)$ فضـای نمونـهٔ آزمـایش X و $P=(p_1,\ldots,p_n)$ توزیـع احتمال نظیر آن باشد. در این صورت داریم

(الف) (۲۲.۱)
$$H(P)\le \log n,$$

تساوی برقرار است اگر و تنها اگر $p_i=\frac{1}{n}$ برای تمام $i=1,\ldots,n$.

(ب) (۲۳.۱)
$$H(P)\ge 0,$$

تساوی برقرار است اگر و تنها اگر kای وجود داشته باشد به قسـمی کـه $p_k=1$ در حالی که برای تمام $i\ne k$، $p_i=0$.

برهان

(الف) در طول این اثبات از نابرابری زیر استفاده خواهد شد (با شکل (۲.۱) مقایسه کنید)

$$\ln a\le a-1. \qquad (۲۴.۱)$$

اکنون $H(P)-\log(n)$ را در نظر بگیرید. داریم

$$H(P)-\log n=-\sum_{i=1}^{n}p_i\log p_i-\log n=-\sum_{i=1}^{n}p_i\{\log p_i+\log n\}$$
$$=\sum_{i=1}^{n}p_i\log\{\frac{1}{p_i n}\},$$

از نابرابری $\ln a\le a-1$ نتیجه می‌شود که

$$\log a=\frac{\ln a}{\ln 2}\le (a-1)\frac{\ln e}{\ln 2}=(a-1)\log e. \qquad (۲۵.۱)$$

با بهره‌گیری از این نابرابری داریم

$$H(P)-\log n\le \sum_{i=1}^{n}p_i\left\{\frac{1}{p_i n}-1\right\}\log e=\left\{\sum_{i=1}^{n}\frac{1}{n}-\sum_{i=1}^{n}p_i\right\}\log e$$
$$=\left\{n\frac{1}{n}-1\right\}\log e=0. \qquad (۲۶.۱)$$

در نتیجه

$$H(P) \le \log n,$$

تساوی برقرار است اگر و تنها اگر $\frac{۱}{(p_i n)} = ۱$ که متناظر با $a = ۱$ در شکل (۲.۱) می‌باشد. این بدین معنی است که $p_i = \frac{۱}{n}$، برای تمام $i = ۱, \ldots, n$.

(ب) چون p_i و $-\log p_i$ هر دو نمی‌توانند منفی باشند، مقدار اطلاع همواره مثبت یا مساوی صفر است. از این رو داریم

$$H(P) \ge ۰.$$

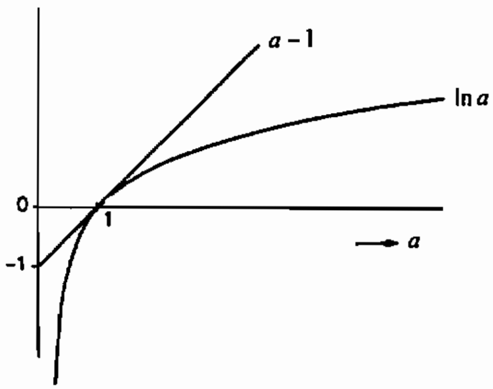
به سهولت دیده می‌شود که $H(P)$ برابر صفر است اگر یک مؤلفهٔ P برابر یک و بقیهٔ احتمالها برابر صفر باشند. ■

مقدار ماکسیمم اطلاع در این صورت برابر $\log n$ است. برای این‌که اثری از مقدار اطلاع یک سیستم به دست آوریم، به مثال زیر توجّه می‌کنیم.

مثال ۲.۱

یک تصویر تلویزیونی شامل ۵۷۶ خط است که هر یک از ۷۲۰ عضو تصویری ساخته شده است. بنابراین یک تصویر تلویزیونی در مجموع شامل ۴۱۴۷۲۰ عضو تصویری است. با این فرض که یک تصویر مدرج خاکستری که در آن هر عضو تصویر می‌تواند یکی از ۱۰ فاصلهٔ شدّت را نمایش دهد تعداد $۱۰^{۴۱۴۷۲۰}$ تصویر تلویزیونی متفاوت امکان‌پذیر است. اگر هر یک از این تصاویر با احتمال مساوی رخ دهد، مقدار اطلاع موجود در یک تصویر برابر است با

$$H(P) = \log n = \log(۱۰^{۴۱۴۷۲۰}) \approx ۱{,}۴ \times ۱۰^{۶} \text{ بیت}.$$ ▲



شکل ۲.۱- تفسیر نموداری $\ln a \le a - ۱$

در بالا چند ویژگی اندازهٔ اطلاع شانون را بررسی کردیم. البته هنوز ویژگیهای دیگری وجود دارند که می‌توان با در نظر گرفتن این اندازهٔ اطلاع به دست آورد. اینها را در فصلهای آتی ملاحظه خواهید کرد.

دیدیم که اگر احتمالها را به ترتیب متفاوتی قرار دهیم مقدار اطلاع تغییر نمی‌کند. اکنون دو توزیع احتمال زیر را در نظر بگیرید

$$P = \{۰٫۵۰, ۰٫۲۵, ۰٫۲۵\}$$

و

$$Q = \{۰٫۴۸, ۰٫۳۲, ۰٫۲۰\}.$$

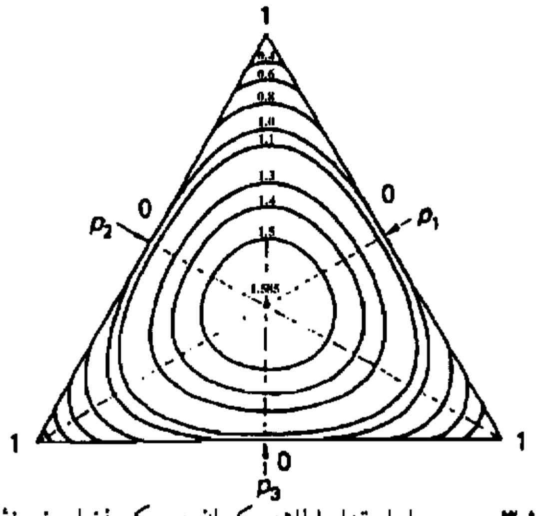
وقتی مقدار اطلاع برای هر دو حالت مربوطه را حساب کنیم، به دست می‌آوریم

$$H(P) = H(Q) = ۱٫۵ \text{ بیت}.$$

یعنی توزیعهای احتمال متفاوت می‌توانند به مقدار اطلاع یکسانی منجر گردند: برخی از آزمایشها می‌توانند توزیعهای احتمال متفاوتی داشته باشند ولی مقدار اطلاع آنها یکسان باشد. شکل (۳.۱) به طور هندسی نشان می‌دهد که هر یک از توزیعهای احتمال منجر به مقدار اطلاع یکسان برای سه مقدار احتمال ($n = ۳$) می‌گردد.

منحنیهای بسته توزیعهای احتمالی را نشان می‌دهند که به مقدار اطلاع یکسانی منجر می‌گردند. مقادیر احتمالهای نظیر برای هر نقطه روی یک منحنی را می‌توان با تصویر بر خطوط $p_۱$، $p_۲$ و $p_۳$ به دست آورد.

به سادگی می‌توان ثابت کرد که ماکسیمم اطلاع برای $n = ۳$ برابر است با



شکل ۳.۱- منحنیها با مقدار اطلاع یکسان در یک فضای نمونهٔ سه‌تایی

$$H(P) = \log ۳ = ۱٫۵۸ \text{ بیت}.$$

چون تنها یک توزیع احتمال وجود دارد که می‌تواند به ماکسیمم مقدار اطلاع منجر گردد، یعنی $P = \{\frac{۱}{۳}, \frac{۱}{۳}, \frac{۱}{۳}\}$، در این حالت به جای یک منحنی بسته دقیقاً یک نقطه در شکل (۳.۱) به دست می‌آید.

برای به دست آوردن بصیرت بیشتری از آنچه که اندازهٔ اطلاع شانون بیان می‌کند، دو مثال زیر را بررسی می‌کنیم.

مثال ۳.۱

فرض کنید منطقه‌ای شامل ۱۶ ناحیه است که یکی از آنها سایه‌دار است (شکل (۴.۱) را ببینید). با پرسش سؤالاتی که تنها می‌توان با بلی و خیر پاسخ داد، می‌خواهیم تعیین کنیم که ناحیهٔ سایه‌دار در کجا قرار دارد. بهترین استراتژی چیست؟ می‌توان حدس زد؛ ولی در این صورت باید این مخاطره را پذیرفت که قبل از این‌که سرانجام ناحیهٔ سایه‌دار را بیابیم ۱۶ سؤال بپرسیم. بهتر است که به طور انتخابی عمل کنیم. در این صورت بازی پرسش و پاسخ می‌تواند به صورت زیر پایان پذیرد، برای مثال (همچنین شکل (۵.۱) را ملاحظه کنید):

۱. آیا ناحیهٔ سایه‌دار یکی از ۸ ناحیهٔ پایین شکل است؟
پاسخ: «خیر»، بنابراین ناحیه‌های ۹ تا ۱۶ را می‌توان حذف کرد.
۲. آیا ناحیهٔ سایه‌دار یکی از ۴ ناحیهٔ باقی‌ماندهٔ سمت چپ است؟
پاسخ: «بلی»، بنابراین ناحیهٔ سایه‌دار ۱، ۲، ۵ یا ۶ است.
۳. آیا ناحیهٔ سایه‌دار یکی از ۲ ناحیه از چهار ناحیهٔ باقی‌ماندهٔ پایین است؟
پاسخ: «بلی»، از این رو ناحیهٔ سایه‌دار ۵ یا ۶ است.
۴. آیا ناحیهٔ سمت چپ است؟
پاسخ: «خیر»، بنابراین ناحیهٔ سایه‌دار ۶ است.

در نتیجه برای تعیین این که کدام یک از ۱۶ ناحیه سایه‌دار است، چهار سؤال لازم است.

اکنون اگر با توجّه به این مسأله مقدار اطلاع را بررسی کنیم، چون تمام ۱۶ ناحیه دارای احتمال مساویند، در می‌یابیم که

$$H(P) = -\sum_{i=۱}^{۱۶} \frac{۱}{۱۶} \log \frac{۱}{۱۶} = \log(۱۶) = ۴ \text{ بیت}.$$

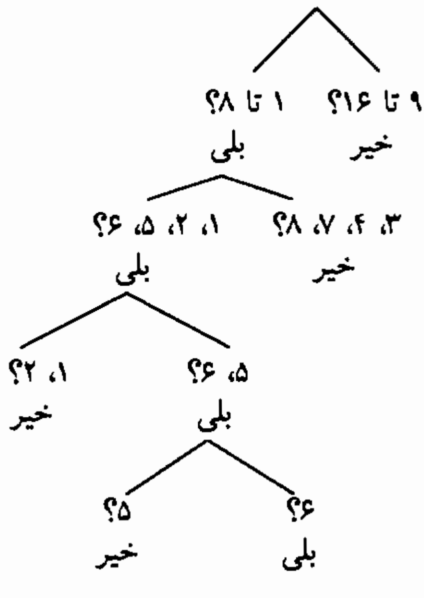
مقدار اطلاع ظاهراً با می‌نیمم تعداد سؤالاتی که بایستی برای تعیین این‌که کدام برآمد (ناحیهٔ سایه‌دار در این حالت) رخ داده است متناظر می‌باشد. ▲

این مطلب را در مثال زیر وقتی توزیعهای احتمالی که در آن همهٔ احتمالها برابر نیستند و تفسیر مثال (۳.۱) نیز برقرار است بررسی می‌کنیم.

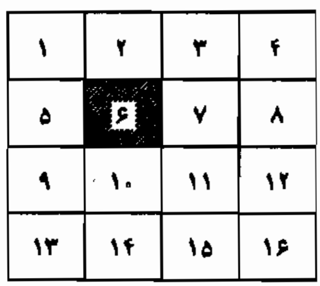
مثال ۴.۱

فضای نمونهٔ X به صورت $X=\{x_1,x_2,x_3\}$ داده شده است. در صورتی که فضای احتمال همراه آن به صورت $P=\{\frac{1}{2},\frac{1}{4},\frac{1}{4}\}$ داده شده باشد، بازی «بلی» و «خیر» را مجدداً اجرا می‌کنیم، بدیهی به نظر می‌رسد که نخست برای x_1 بپرسیم؛ چون این برآمد بیشترین احتمال را دارد.

اگر پاسخ، «بلی» است، در این صورت برآمد را در یک مرحله پیدا نموده‌ایم. اگر



شکل ۵.۱- مثال ساختار درختی برای بازی پرسش و پاسخ



شکل ۴.۱- بازی پرسش و پاسخ: یافتن ناحیهٔ سایه‌دار

پاسخ «خیر» باشد، در این صورت به طور وضوح برآمد x_2 یا x_3 است. برای تعیین x_2 یا x_3 به سؤال دیگری نیاز است. بنابراین به طور کلی برای دانستن برآمد، نیاز به دو سؤال است. بنابراین بایستی یک سؤال یا دو سؤال با احتمال مساوی بپرسیم، از این رو متوسط

آن ۱٫۵ سؤال می‌باشد.

اگر مقدار اطلاع را طبق اندازهٔ شانون محاسبه کنیم، در این صورت داریم

$$H(P) = -\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۴}\log\frac{۱}{۴} - \frac{۱}{۴}\log\frac{۱}{۴} = ۱٫۵ \text{ بیت}.$$

در نتیجه تفسیر بیان شدهٔ قبلی برای احتمالهای نابرابر نیز برقرار است. ▲

۴.۱ اندازهٔ اطلاع شرطی، توأم و متقابل

در بخش (۲.۱) به یک آزمایش احتمالی (X,Y) با برآمدهای ممکن (x_i, y_j) که در آن $(x_i, y_j) \in (X,Y)$ اشاره نمودیم.

براساس حجم فضای نمونه (X,Y) می‌توان نتیجه گرفت که آزمایش (X,Y) در مجموع دارای nm برآمد توأم ممکن می‌باشد. اکنون اگر بخواهیم مقدار اطلاع مربوط به (X,Y) را تعیین کنیم، از روش زیر استفاده می‌کنیم.

تعداد nm پیشامد توأم (x_i, y_j) با احتمال رخداد $r(x_i, y_j)$ یا r_{ij} وجود دارد (بخش (۲.۱) را برای نماد ببینید). حالا فرض کنید که nm پیشامد توأم را به صورت پیشامدهای z_1، z_2، ...، z_{nm} و احتمال متناظر با آن را به صورت $p(z_1)$، $p(z_2)$، ...، $p(z_{nm})$ نشان دهیم. در این صورت در واقع مجدداً یک فضای نمونهٔ یک‌بعدی داریم، و بنا بر تعریف اندازهٔ اطلاع حاشیه‌ای داریم

$$H(Z) = -\sum_{k=1}^{nm} p(z_k)\log p(z_k). \qquad (۲۷.۱)$$

لکن چون هر $p(z_k)$ برابر با یکی از احتمالهای $r(x_i, y_j)$ خواهد بود، مجموع روی k با مجموع روی i و j یکی خواهد بود. به عبارت دیگر:

$$H(Z) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i, y_j)\log[r(x_i, y_j)].$$

این مطلب تعریف اندازهٔ اطلاع توأم زیر را نتیجه می‌دهد.

تعریف ۲.۱

یک آزمایش احتمالی (X,Y) با فضای نمونهٔ دو بعدی (X,Y) که در آن r_{ij} یا $r(x_i, y_j)$ احتمال x_i و y_j می‌باشد در نظر بگیرید. در این صورت ***اندازهٔ اطلاع توأم*** را به

صورت زیر تعریف می‌کنیم:

$$H(X,Y)=-\sum_{i=1}^{n}\sum_{j=1}^{m}r(x_i,y_j)\log[r(x_i,y_j)] \qquad (۲۸.۱)$$

نمادهای دیگری مانند $H(R)$ و $H(r_{11},\dots,r_{nm})$ را علاوه بر $H(X,Y)$ بـه طـور قـابل تعویضی به کار خواهیم برد.

تا کنون دیدیم که اندازهٔ اطلاع حاشیه‌ای را می‌توان بر مبنـای احتمالهـای حاشـیه‌ای تعریف کرد و احتمالهای توأم منجر به معرّفی اندازهٔ اطلاع توأم می‌گردد. اکنـون بررسـی خواهیم کرد که آیا اندازهٔ اطلاع شرطی را می‌توان برحسب احتمالهای شرطی تعریف کرد.

آزمایشهای احتمالی X و Y را مجدداً در نظر می‌گـیریم. اکنـون فـرض کنیـد کـه علاقه‌مند به مقدار اطلاع نسبت به Y تحت این شرط باشیم که برآمد x_i قبلاً رخ داده است. در این صورت به جای احتمالهای $q(y_j)$، $j=1,\dots,m$، احتمالهای $q(y_j|x_i)$، $j=1,\dots,m$ را داریم ولی هنوز هم مجموع برابر یک است.

در این صورت مقدار اطلاع نسبت به Y به شرط معلوم بودن x_i را می‌توان با مقایسه با اندازهٔ اطلاع حاشیه‌ای به صورت زیر تعریف کرد

$$H(Y|x_i)=-\sum_{j=1}^{m}q(y_j|x_i)\log[q(y_j|x_i)]. \qquad (۲۹.۱)$$

اکنون با محاسبهٔ متوسط روی تمام مقادیر x_i، متوسط مقدار اطلاع Y با معلوم بودن X به صورت زیر به دست می‌آید:

$$\sum_{i=1}^{n}p(x_i)H(Y|x_i)=\sum_{i=1}^{n}p(x_i)\left\{-\sum_{j=1}^{m}q(y_j|x_i)\log[q(y_j|x_i)]\right\}$$

$$=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)q(y_j|x_i)\log[q(y_j|x_i)]$$

$$=-\sum_{i=1}^{n}\sum_{j=1}^{m}r(x_i,y_j)\log[q(y_j|x_i)].$$

این کمیّت با مقدار اطلاع شرطی $H(Y|X)$ نشان داده می‌شود و تعریف زیر حاصل می‌گردد.

تعریف ۳.۱

اندازهٔ اطلاع شرطی نسبت به آزمایش Y به شرط X برابر است با

$$H(Y|X)=-\sum_{i=1}^{n}\sum_{j=1}^{m}r(x_i,y_j)\log[q(y_j|x_i)]. \qquad (۳۰.۱)$$

با روشی مشابه می‌توان مقدار اطلاعی را که به طور متوسط برای X وقتی Y معلوم باشد به دست می‌آید به صورت زیر تعریف کرد

$$H(\mathrm{X}|\mathrm{Y}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i, y_j)\log[p(x_i|y_j)]. \qquad (۳۱.۱)$$

به جای $H(\mathrm{Y}|\mathrm{X})$ و $H(\mathrm{X}|\mathrm{Y})$ نمادهای وابسته به فضای آنها: $H(Y|X)$ و $H(X|Y)$ را نیز به کار خواهیم برد.

قضیهٔ زیر مقدار می‌نیمم و ماکسیمم $H(\mathrm{Y}|\mathrm{X})$ را می‌دهد.

قضیهٔ ۲.۱

فرض کنید $H(\mathrm{Y}|\mathrm{X})$ اندازهٔ اطلاع Y به شرط X باشد. در این صورت داریم

(الف) $$H(\mathrm{Y}|\mathrm{X}) \geq 0, \qquad (۳۲.۱)$$

(ب) $$H(\mathrm{X}|\mathrm{Y}) \leq H(\mathrm{Y}), \qquad (۳۳.۱)$$

اگر X و Y به طور احتمالی مستقل باشند تساوی برقرار است.

برهان

(الف) چون برای تمام i و jها داریم $q(y_j|x_i) \leq 1$، نتیجه می‌شود که $\{-\log p(y_j|x_i)\} \geq 0$. بنابراین مستقیماً از تعریف نتیجه می‌شود که

$$H(\mathrm{Y}|\mathrm{X}) \geq 0.$$

(ب)

$$H(\mathrm{Y}|\mathrm{X}) - H(\mathrm{Y}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i, y_j)\log[q(y_j|x_i)] + \sum_{j=1}^{m} q(y_j)\log q(y_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i, y_j)\log\left[\frac{q(y_j)}{q(y_j|x_i)}\right].$$

با استفاده از نابرابری $\ln(a) \leq a - 1$ (شکل (۲.۱) را ببینید) که قبلاً یادآور شدیم نتیجه می‌شود که

$$H(\mathrm{Y}|\mathrm{X}) - H(\mathrm{Y}) \leq \sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i, y_j)\left[\frac{q(y_j)}{q(y_j|x_i)} - 1\right]\log e.$$

سمت راست این نابرابری را می‌توان به صورت زیر نوشت

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i)q(y_j|x_i)\frac{q(y_j)}{q(y_j|x_i)}\log e-\sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i,y_j)\log e$$

$$=\log e-\log e=0.$$

بنابراین

$$H(\mathrm{Y}|\mathrm{X})\le H(\mathrm{Y}).$$

اگر برای تمام iها و jها، $q(y_j)=q(y_j|x_i)$ دو مقدار اطلاع با هم برابرند و این در حالت استقلال درست است. ■

نتیجه‌ای که می‌توان از این قضیه گرفت این است که مقدار شرطی اطلاع همواره کوچکتر یا مساوی مقدار حاشیه‌ای اطلاع است. به عبارت دیگر، به طور کلّی اطلاع در مورد X موجب کاهش عدم‌حتمیت می‌شود. این مطلب مطابق با ایدهٔ ذهنی دربارهٔ آگاهی قبلی است.

رابطهٔ مستقیمی بین اندازهٔ اطلاع حاشیه‌ای، شرطی و توأم وجود دارد که در قضیهٔ زیر بیان می‌شود.

قضیهٔ ۳.۱

برای تمام آزمایشهای X و Y، داریم:

$$\begin{aligned}H(\mathrm{X},\mathrm{Y})&=H(\mathrm{X})+H(\mathrm{Y}|\mathrm{X})\\&=H(\mathrm{Y})+H(\mathrm{X}|\mathrm{Y}).\end{aligned}\tag{۳۴.۱}$$

برهان

داریم

$$H(\mathrm{X},\mathrm{Y})=-\sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i,y_j)\log[p(x_i)q(y_i|(x_i)]$$

$$=-\sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i,y_j)\log p(x_i)-\sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i,y_j)\log[q(y_j|x_i)]$$

$$=H(\mathrm{X})+H(\mathrm{Y}|\mathrm{X}).$$

اثبات $H(\mathrm{X},\mathrm{Y})=H(\mathrm{Y})+H(\mathrm{X}|\mathrm{Y})$ به روش مشابه انجام می‌شود. ■

آنچه که قضیه واقعاً بیان می کند این است که مقدار اطلاع توأم برابر مجموع مقدار اطلاع نسبت به X و مقدار اطلاع Y به شرط X می‌باشد.

بنابر قضیهٔ (۲.۱) و قضیهٔ (۳.۱) می‌توان نتیجهٔ زیر را به دست آورد

$$H(\mathrm{X},\mathrm{Y}) = H(\mathrm{X}) + H(\mathrm{Y}|\mathrm{X}) \le H(\mathrm{X}) + H(\mathrm{Y}). \qquad (۳۵.۱)$$

و اگر X و Y مستقل باشند تساوی برقرار است.

از این رو می‌توان فرض کرد که اگر دو آزمایش احتمالی مستقل باشند مقدار احتمال توأم ماکسیمم و وقتی وابستگی افزایش می‌یابد کاهش پیدا می‌کند. با وابستگی مطلق، برآمد Y معلوم است اگر برآمد X معلوم باشد، بنابراین داریم $H(\mathrm{Y}|\mathrm{X}) = ۰$. در این حالت $H(\mathrm{X},\mathrm{Y}) = H(\mathrm{X})$.

اکنون آخرین تعریف را در این بخش بیان می‌کنیم که در مورد اندازهٔ اطلاع متقابل است و نقش مهمی را نسبت به ظرفیت کانال ارتباطی بازی می‌کند و بعداً بررسی می‌شود.

تعریف ۴.۱

اندازهٔ اطلاع متقابل مربوط به X و Y به صورت زیر تعریف می‌شود:

$$\begin{aligned} I(\mathrm{X};\mathrm{Y}) &= H(\mathrm{Y}) - H(\mathrm{Y}|\mathrm{X}) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} r(x_i,y_j)\log\left[\frac{r(x_i,y_j)}{p(x_i)q(y_j)}\right]. \end{aligned} \qquad (۳۶.۱)$$

$I(\mathrm{X};\mathrm{Y})$ را می‌توان به عنوان اندازه‌ای برای وابستگی بین X و Y تفسیر نمود. وقتی X و Y مستقلند، $I(\mathrm{X};\mathrm{Y})$ می‌نیمم است، یعنی

$$I(\mathrm{X};\mathrm{Y}) = ۰$$

اگر Y کاملاً وابسته به X باشد، آن‌گاه $H(\mathrm{Y}|\mathrm{X}) = ۰$ و $I(\mathrm{X};\mathrm{Y})$ ماکسیمم مقدار خودش را به دست می‌آورد که برابر است با

$$I(\mathrm{X};\mathrm{Y}) = H(\mathrm{Y}).$$

به خواننده واگذار می‌شود که نشان دهد برای تمام Xها و Yها داریم

$$\begin{aligned} I(\mathrm{X};\mathrm{Y}) &= H(\mathrm{X}) - H(\mathrm{X}|\mathrm{Y}) \\ &= H(\mathrm{X}) + H(\mathrm{Y}) - H(\mathrm{X},\mathrm{Y}), \end{aligned} \qquad (۳۷.۱)$$

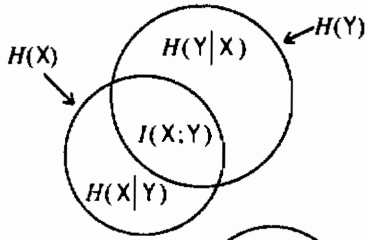
و $I(\mathrm{X};\mathrm{Y})$ متقارن است؛ یعنی برای تمام Xها و Yها داریم

$$I(\mathrm{X};\mathrm{Y}) = I(\mathrm{Y};\mathrm{X}). \qquad (۳۸.۱)$$

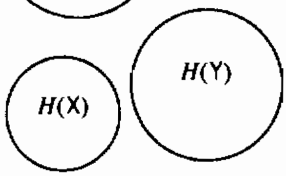
در این بخش سه اندازهٔ اطلاع، یعنی اندازه‌های اطلاع شرطی، توأم و متقابل را تعریف کردیم. توجّهی نیز به روابط مختلف بین اندازه‌ها شد. این مطالب توسط نمودار وِن در

شکل (۶.۱) به بهترین وجهی شرح داده شده و خلاصه شده‌اند.

(الف)



(ب)



شکل ۶.۱- روابط بین اندازه‌های اطلاع: (الف) حالت کلّی، (ب) حالت استقلال

داریم

$$I(\mathsf{X};\mathsf{Y}) = H(\mathsf{X}) \cap H(\mathsf{Y}),$$

$$H(\mathsf{X},\mathsf{Y}) = H(\mathsf{X}) \cup H(\mathsf{Y}).$$

از شکل (۶.۱- الف)، حالت کلّی، می‌توان نتیجه گرفت که

- $H(\mathsf{X}|\mathsf{Y}) \le H(\mathsf{X})$ و $H(\mathsf{Y}|\mathsf{X}) \le H(\mathsf{Y})$;
- $I(\mathsf{X};\mathsf{Y}) \le H(\mathsf{Y})$ و $I(\mathsf{X};\mathsf{Y}) \le H(\mathsf{X})$;
- $I(\mathsf{X};\mathsf{Y}) = H(\mathsf{X}) - H(\mathsf{X}|\mathsf{Y}) = H(\mathsf{Y}) - H(\mathsf{Y}|\mathsf{X});$
- $$\begin{aligned} H(\mathsf{X},\mathsf{Y}) &= H(\mathsf{X}|\mathsf{Y}) + I(\mathsf{X};\mathsf{Y}) + H(\mathsf{Y}|\mathsf{X}) \\ &= H(\mathsf{Y}) + H(\mathsf{X}|\mathsf{Y}) = H(\mathsf{X}) + H(\mathsf{Y}|\mathsf{X}); \end{aligned}$$
- $H(\mathsf{X},\mathsf{Y}) \le H(\mathsf{X}) + H(\mathsf{Y}).$

در شکل (۶.۱- ب)، X و Y مستقلّند. اکنون توجّه کنید که

- $I(\mathsf{X};\mathsf{Y}) = \circ;$
- $H(\mathsf{X},\mathsf{Y}) = H(\mathsf{X}) + H(\mathsf{Y});$
- $H(\mathsf{X}) = H(\mathsf{X}|\mathsf{Y})$ و $H(\mathsf{Y}) = H(\mathsf{Y}|\mathsf{X}).$

روابطی را که بین اندازه‌های اطلاع گوناگون در این بخش و بخشهای قبلی به دست آوردیم به سادگی می‌توان با نمودارهای وِن نمایش داد.

۵.۱ اصول موضوعه

در بخش (۳.۱) اندازهٔ اطلاع شانون معرّفی شد و برخی از ویژگیهای اندازهٔ اطلاع را به دست آوردیم. به نظر می‌رسد این ویژگیها متناظر با ویژگیهایی است که با درک مستقیم از یک اندازهٔ اطلاع انتظار خواهیم داشت. در حالت یک‌نواخت اندازهٔ اطلاع شانون با اندازهٔ اطلاع هارتلی که لگاریتم تعداد پیامهاست برابر می‌باشد (برابری (۱.۱) را مقایسه کنید). اندازهٔ اطلاع شانون براساس احتمالهایی است که می‌توان مستقیماً از حالت یک‌نواخت و اندازهٔ هارتلی به دست آورد.

فرض کنید که یک اندازهٔ اطلاع بایستی در سه شرط زیر صدق کند.

(الف) اگر تمام برآمدها به گروههایی تقسیم شوند، در این صورت مقادیر H برای همهٔ گروههای مختلف ضرب در وزن آماری آنها باید برابر کل H باشد؛

(ب) H بایستی برحسب p_i پیوسته باشد؛

(پ) اگر همهٔ p_iها برابر باشند، یعنی برای همهٔ iها، $p_i = \frac{۱}{n}$، در این صورت H به عنوان تابعی از n به طور یک‌نواخت افزایش خواهد یافت. این بدین معنی است که عدم‌حتمیت با افزایش تعداد احتمالهای برابر افزایش خواهد یافت.

برای n برآمد با احتمال برابر، H بایستی طبق اندازهٔ هارتلی و شرط (پ) در $H = \log n$ صدق کند. برای احتمالهای نابرابر حالت زیر را بررسی می‌کنیم. فرض کنید احتمالها برابر $\frac{۳}{۶}$، $\frac{۲}{۶}$ و $\frac{۱}{۶}$ باشند. شکل (۷.۱- الف) درخت تصمیمی را می‌دهد که باید H را برای آن حساب کرد.

مقدار H نسبت به درخت تصمیم شکل (۷.۱- پ) باید برابر با $H_{پ} = \log ۶$ باشد (با شرط (پ) مقایسه کنید). با وجود این، چون هر دو درخت تصمیم شکل (۷.۱- ب) و (۷.۱- پ) اصولاً با هم برابرند، مقدار H نسبت به درخت تصمیم (۷.۱- ب) نیز باید برابر $\log ۶$ باشد؛ $H_{ب} = \log ۶$. بر مبنای شرط (الف) این مقدار باید برابر باشد با عدم‌حتمیت با توجّه به انتخاب بین شاخه‌هایی که با $\frac{۳}{۶}$، $\frac{۲}{۶}$ و $\frac{۱}{۶}$ در شکل (۷.۱- ب) نشان داده شده‌اند (یعنی، جست‌وجو کردن $H_{الف}$) به اضافهٔ عدم‌حتمیتها نسبت به زیر شاخه ضرب در وزن

آنها.

داریم

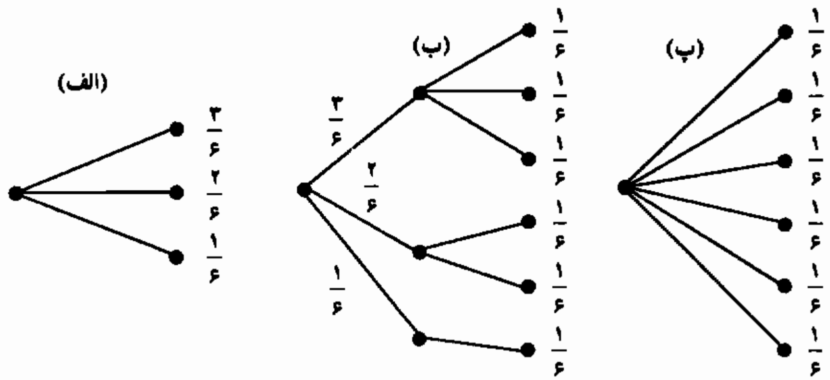
$$H_{\text{الف}} + \frac{۳}{۶}\log ۳ + \frac{۲}{۶}\log ۲ + \frac{۱}{۶}\log ۱ = \log ۶$$

و

$$H_{\text{الف}} = -\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۳}\log\frac{۱}{۳} - \frac{۱}{۶}\log\frac{۱}{۶}$$

به طور کلّی نتیجه می‌شود که

$$H(\mathrm{X}) = -\sum_{i=۱}^{n} p_i \log p_i$$



شکل ۷.۱- درختهای تصمیم مربوط به برآمدهایی با احتمال نابرابر

چاندی[1] و مک‌لوید[2] (۱۹۶۰) قضیهٔ توصیفی زیر را که به طور یکتا اندازهٔ اطلاع شانون را تعیین می‌کند ارائه نمودند.

قضیهٔ ۴.۱

تابع $f(\mathrm{X}) = f(P) = f(p_۱,\dots,p_n)$ و تابع $g(.)$ را که در ویژگیهای زیر صدق می‌کنند در نظر بگیرید:

(الف) $f(P) = \sum_{i=۱}^{n} g(p_i)$،

(ب) $f(.)$ در فاصلهٔ $[۰,۱]$ پیوسته است،

1. Chaundy 2. McLeod

(پ) $f(P)$ جمع‌پذیر است،

$$f(p_1q_1,\dots,p_nq_m)=f(p_1,\dots,p_n)+f(q_1,\dots,q_m)$$

(ت) $f(\frac{1}{2},\frac{1}{2})=1$.

در این صورت

$$f(P)=H(P)=-\sum_{i=1}^{n}p_i\log p_i\,. \qquad \blacksquare$$

از قضیه نتیجه می‌شود که در واقع ویژگی جمع‌پذیری است که اندازهٔ اطلاع شانون را به طور یکتا تعیین می‌کند.

گذشته از این توجّه کنید که چون شانون اندازهٔ اطلاعش را در ۱۹۴۸ معرّفی کرد، پژوهشهای متنوّعی برای یافتن شقّ دیگری از اندازهٔ اطلاع شانون انجام شده است. در این جا به خصوص باید از اثر رنی[1] (۱۹۶۰)، داروچی[2] (۱۹۷۰) و آزیموتو[3] (۱۹۷۱) یاد کرد. با توجّه به دو اثر اخیر شرط قوی جمع‌پذیری با شکل ضعیفتری از جمع‌پذیری جایگزین شده است. وَن درِ لُوب (۱۹۸۱) همهٔ این اندازه‌ها را با یک قالب متّحدالشکل بیان کرد.

۶.۱ الگوی ارتباطی

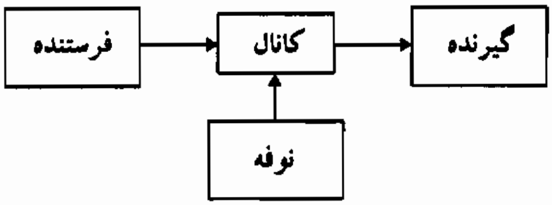
معمولاً اطلاع یک منبع به خودی خود بیش از این مورد استفاده قرار نمی‌گیرد. به دلایل تاریخی، معمول این است که از روشی که در آن اطلاع برحسب ***الگوی ارتباطی*** به کار برده می‌شود صحبت شود. در حالت الگوی ارتباطی، تأکید بر انتقال اطلاع همان طور که توسط منبع تولید می‌شود به یک مقصد می‌باشد. ذخیرهٔ اطلاع در یک حافظه امروزه همچنان دارای اهمیت زیاد می‌باشد و گرچه یک مسألهٔ ارسال نمی‌باشد می‌توان ذخیره را بر حسب آن تشریح کرد.

در حین انتقال اطلاع بین منبعی که اطلاع تولید می‌کند که آن را اغلب ارسال‌کننده یا فرستنده می‌نامیم از یک طرف و مقصد یا گیرنده از طرف دیگر ارتباط برقرار می‌شود. الگوی اساسی در شکل (۸.۱) ترسیم شده است. مسألهٔ اصلی در ارتباطات بین فرستنده و گیرنده خطاها یا تغییر شکلهایی است که می‌تواند در حین انتقال از طریق کانال ارتباطی به عنوان نتیجه نوفه‌ای که بر کانال اثر می‌کند رخ دهد. انتقال اطلاع بایستی تا درجهٔ معیّنی بدون-خطا باشد که بستگی به شرایط تحمیل شده توسط گیرنده دارد. بنابراین بایستی تصحیح

1. Renyi 2. Daroczy 3. Azimoto

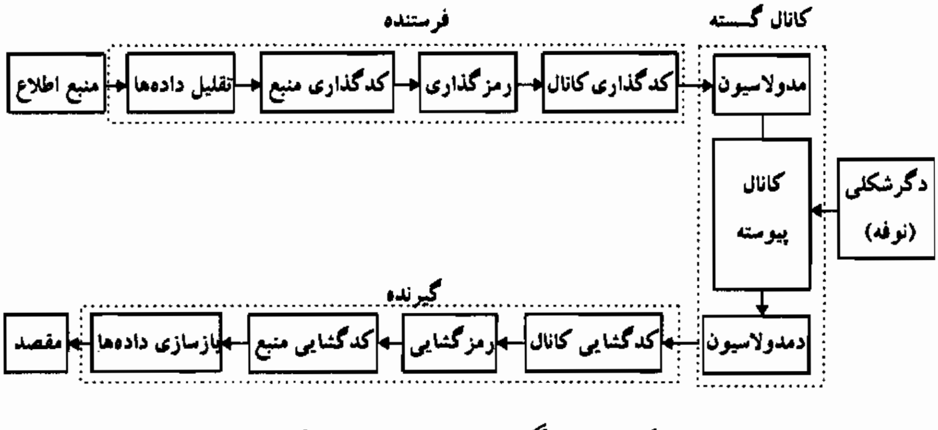
خطا امکان‌پذیر باشد یا انتقال بایستی به قدر کافی خوب باشد به طوری که از بعضی خطاهایی که درجهٔ اهمیت کمتری دارند بتوان صرف‌نظر کرد.

وقتی سیگنالهایی نظیر سخنرانی، موزیک یا ویدئو ارسال می‌کنیم، یک ارسال کامل، یعنی بدون-خطا واقعاً امکان‌پذیر نمی‌باشد و تنها قادر خواهیم بود به مقداری که سیگنال دریافت شده از سیگنال ارسال شده متفاوت است شرایطی را تحمیل کنیم. کیفیت مورد نیاز به انتخاب وسیلهٔ ارسال مناسبی منجر می‌شود، لکن به ویژه شرایط حدّی در تطابق این کانال به فرستنده و گیرنده تحمیل می‌گردد.



شکل ۱.۸- الگوی ارتباطی مقدّماتی

شرح مفصّلتری از الگوی ارتباطی در شکل (۹.۱) ارائه شده است. سیستم ارتباطی بایستی اطلاع تولید شده توسط منبع را در حدّ امکان به طور دقیق به مقصد ارسال نماید. فرض می‌کنیم که منبع اطلاع و مقصد و همچنین کانال همگی معلومند. منبع نوفه‌ای که بر کانال اثر می‌کند نیز معلوم در نظر گرفته می‌شود.



شکل ۹.۱- الگوی ارتباطی با جزئیات کامل

فرض می‌کنیم که کانال پیوسته سیگنالهایی ارسال می‌کند که طبیعت آنها به ارسال فیزیکی موجود یا وسیلهٔ ذخیرهٔ (الکتریکی، مغناطیسی، نوری) و روش انتخابی مدولاسیون

دارد. مشخّصه‌های فیزیکی یک کانال پیوسته مانند پهنای باند و نسبت سیگنال به نوفه را نیز معلوم در نظر می‌گیریم. هدف فرستنده آن است که اطلاع را از منبع اطلاع برای ارسال از طریق کانال ارتباطات به صورت مناسب در آورد، در حالی که گیرنده می‌کوشد که دگرشکلی و خطاهایی را که در کانال به وجود می‌آید تصحیح کند و سپس اطلاع را به فرم مناسبی برای مقصد تبدیل کند.

یک بخش فرعی از توابع فرستنده منجر به چهار وجه فرعی می‌شود. قبل از هر چیز آشکار خواهد شد که تمام اطلاعات تولید شده توسط منبع اطلاع مناسب برای مقصد نمی‌باشند. به علّت ملاحظات کارایی، بهتر آن است که حذف شوند. این کار را ***کاهش داده‌ها*** می‌نامند. اطلاعات باقی‌مانده را ***اطلاع کارا*** می‌نامیم.

اغلب این اطلاع کارا باید با روش دیگری (عددی) مثلاً به شکل دودویی پردازش شود و هنوز هم ساختار ذاتی بسیاری را شامل خواهد بود. به واسطهٔ کاربرد ***کدگذاری منبع***، که گاهی اوقات نیز ***تراکم داده‌ها*** نامیده می‌شود، اطلاع کارا در حدّ امکان متراکم نشان داده می‌شود.

اغلب بیشتر و بیشتر ثابت می‌شود که مطلوب آن است که اطلاع به دست آمده را حفاظت کنیم تا از استفادهٔ غیرواقعی ممکن جلوگیری شود. یک راه حل این است که اطلاعات را با کمک کدهای سری رمزی کنیم.

حراست از اطلاعات در مقابل خطاهای ممکن که می‌تواند در کانال رخ دهد به عنوان عنصر چهارم فرستنده پیش می‌آید. برای رسیدن به این مطلب، اطلاع اضافی که بعداً می‌تواند برای نوسازی اطلاع اصلی، اگر خطایی رخ داده باشد، به کار برده شود اضافه می‌کنیم. وقتی کدی را برای تشخیص و/یا تصحیح خطا به کار می‌بریم آن را ***کدگذاری کانال*** می‌نامیم.

اطلاعی را که از فرستندهٔ این چنین به دست آمده است بعداً آن را به کانال تحویل می‌دهیم. اگر کانال را در سطحی خلاصه کنیم که در آن اطلاع پیشنهاد شده در سمت ورودی به خوبی تفکیک شده باشند و بعد از انتقال نیز نمادها در سمت خروجی مجدداً تولید شوند آن را ***کانال گسسته*** می‌نامیم. در واقع، این ارسال توسط سیگنالهایی که باید از طرق ابزار فیزیکی (کانال پیوسته) فرستاده شوند انجام شود. تبدیل نمادهای پیشنهاد شده به سیگنالهای مناسب توسط مدولاسیون انجام می‌شود. این سیگنالها در الگوی ارتباطی تحت بررسی با نوفه تغییر شکل می‌دهند. از این رو مخلوط سیگنال و نوفه حاصل سرانجام مجدداً با کشف (دمدولاسیون) به نمادها برگشت داده می‌شوند. با وجود این هر نوفه

موجود به طور تصادفی می‌تواند همان نتیجه‌ای را داشته باشد که نماد معرّفی شده در نماد نادرست دیگری پس از ارسال نتیجه می‌دهد. اکنون اطلاع سرچشمه گرفته از کانال برای خطاها و احتمالاً تصحیح شده توسط کاربرد کانال ***کدگشایی*** بررسی می‌شود.

اطلاعات حاصل متوالیاً ***رمزگشایی*** شده و در ***کدگشا*** کدگشایی می‌شود. سرانجام اطلاعات را برای مقصد به وسیلهٔ ***بازسازی داده‌ها*** به شکل دلخواه درمی‌آوریم.

تا حدّ زیادی می‌توان مراحل پردازش مذکور در بالا را به عنوان ارسالهای معیّن در نظر گرفت به این مفهوم که ارسالهای پیش‌رو و پس‌رو مجدداً نتیجهٔ اصلی را به طور دقیق ارائه می‌کنند. استثناها عبارتند از تقلیل داده‌ها و نوسازی داده‌ها و کانال گسسته که فرض شده است نوفه تصادفی در آن ظاهر می‌شود.

در فصلهای آتی این جنبه‌های کانال ارتباطی را بیشتر مورد بررسی قرار خواهیم داد و حدودی که در آن بایستی کارکرد برای این‌که به ارسال کارا و بی‌-خطا دست یابیم یا ذخیرهٔ اطلاع اشاره خواهد شد.

۷.۱ تمرینها

۱.۱ مجموع وجوه دو تاس نرمال پس از پرتاب برابر ۷ است. آیا این امر چقدر اطلاع برای ما فراهم می‌کند؟ پاسخ خود را شرح دهید.

تذکر: برآمدهایی نظیر (۶،۱) و (۱،۶) متفاوت در نظر گرفته می‌شوند.

۲.۱ یک گلدان شامل m توپ سیاه و $n-m$ توپ سفید است. آزمایش X بیرون آوردن تصادفی یک توپ بدون جای‌گذاری در گلدان می‌باشد. آزمایش Y بیرون آوردن تصادفی توپ دوم می‌باشد.

(الف) مقدار اطلاع دریافتی از آزمایش X را تعیین کنید؛

(ب) مقدار اطلاع نسبت به آزمایش Y اگر رنگ توپ انتخاب شده در آزمایش X معلوم نباشد را تعیین کنید؛

(پ) به سؤال (ب) اکنون با این فرض که رنگ توپ انتخاب شده در آزمایش X معلوم است پاسخ دهید.

۳.۱ یک چرخ رولت به ۳۸ بخش شماره‌گذاری شده با رنگهای مختلف تقسیم شده است. توزیع بخشها برطبق رنگ عبارتند از:

۲ سبز،
۱۸ قرمز،
۱۸ سیاه.

آزمایش عبارت است از پرتاب یک توپ کوچک روی چرخ رولت در حال گردش. پیشامد این که توپ در یکی از ۳۸ بخش قرار گیرد دارای احتمالِ مساوی برای هر بخش است.

(الف) آیا چه مقدار اطلاع دریافت می‌کنیم اگر تنها رنگ مورد توجّه باشد؟
(ب) آیا چه مقدار اطلاع دریافت می‌کنیم اگر رنگ و شماره مورد توجّه باشد؟
(پ) در این صورت برای اطلاع شرطی اگر رنگ معلوم باشد چه نتیجه‌ای به دست می‌آید؟

۴.۱ آوندی شامل ۵ توپ سیاه و ۱۰ توپ سفید است. آزمایش X بیرون آوردن تصادفی یک توپ می‌باشد. آزمایش Y بیرون آوردن تصادفی یک توپ است وقتی توپ استخراجی در آزمایش X به آوند باز گردانده نشود. رنگ توپ استخراجی مورد توجّه است.

(الف) آیا آزمایش X شامل چه‌قدر عدم‌حتمیت می‌باشد؟
(ب) عدم‌حتمیت در آزمایش Y به شرط آن که توپ اوّل سیاه باشد چه‌قدر است؟
(پ) عدم‌حتمیت در آزمایش Y به شرط آن که توپ اوّل سفید باشد چه‌قدر است؟
(ت) آزمایش Y شامل چه مقدار عدم‌حتمیت است؟

۵.۱ برای امتحان معیّنی، ۷۵٪ شرکت‌کنندگان قبول می‌شوند و ۲۵٪ قبول نمی‌شوند. از دانشجویانی که قبول شده‌اند ۱۰٪ اتومبیل دارند و از آنهایی که رد شده‌اند ۵۰٪ اتومبیل دارند.

(الف) اگر نتیجهٔ امتحان یک دانشجو گفته شود چه مقدار اطلاع دریافت می‌کنیم؟
(ب) در اعلان این که دانشجویی که قبول شده دارای اتومبیل هست یا نیست چه مقدار اطلاع وجود دارد؟
(پ) اگر نتیجه امتحان یک دانشجوی اتومبیل‌دار را اعلان کنند چه مقدار عدم‌حتمیت باقی می‌ماند؟

۶.۱ در ناحیهٔ مشخّصی ۲۵٪ دختران بور هستند و ۷۵٪ از تمام دختران بور، چشم آبی هستند. در هر یک از حالات زیر چه مقدار اطلاع کسب می‌کنیم؟

(الف) اگر بدانیم که یک دختر بور است و رنگ (آبی یا غیرآبی) چشمش را به ما گفته باشند؛

(ب) اگر بدانیم که یک دختر چشم آبی است و رنگ (بور یا غیربور) مویش را به ما گفته باشند؛

(پ) اگر هر دو رنگ مو و چشمش را به ما گفته باشند.

۷.۱ از یک گروه دانشجویی، ۲۵٪ برای ورود به دانشگاه مناسب نیستند. با وجود این به عنوان نتیجهٔ یک گزینش، تنها ۷۵٪ از این دانشجویان نامناسب رد شده‌اند. ۵۰٪ از تمام دانشجویان رد شده‌اند.

(الف) اگر دانشجویی که می‌داند برای دانشگاه مناسب نیست نتیجهٔ گزینش را بشنود چه مقدار اطلاع دریافت می‌کنیم.

(ب) اگر انتخاب با پرتاب سکّه معلوم شود، سؤال مشابه (الف) را پاسخ دهید.

(پ) نتایج (الف) و (ب) را مقایسه کنید و توصیفی برای اختلافها بیان کنید.

۸.۱ دو آزمایش X و Y داده شده‌اند. فضای نمونه نسبت به X شامل $x_۱$، $x_۲$ و $x_۳$ و مربوط به Y شامل $y_۱$، $y_۲$ و $y_۳$ می‌باشد. احتمالهای توأم $r(x_i, y_j) = r_{ij}$ در ماتریس R زیر داده شده است.

$$\begin{bmatrix} r_{۱۱} & r_{۱۲} & r_{۱۳} \\ r_{۲۱} & r_{۲۲} & r_{۲۳} \\ r_{۳۱} & r_{۳۲} & r_{۳۳} \end{bmatrix} = \begin{bmatrix} \frac{۷}{۲۴} & \frac{۱}{۲۴} & ۰ \\ \frac{۱}{۲۴} & \frac{۱}{۴} & \frac{۱}{۲۴} \\ ۰ & \frac{۱}{۲۴} & \frac{۷}{۲۴} \end{bmatrix}$$

(الف) اگر فردی نتیجهٔ برآمد X و Y را به شما بگوید چه مقدار اطلاع دریافت می‌کنید؟

(ب) اگر شخصی نتیجهٔ برآمد Y را به شما بگوید چه مقدار اطلاع کسب می‌کنید؟

(پ) اگر شخصی نتیجهٔ برآمد X را به شما بگوید در حالی که شما قبلاً نتیجهٔ برآمد Y را می‌دانستید چه مقدار اطلاع دریافت می‌کنید؟

۹.۱ یک سیستم ارتباطی دودویی از نمادهای «صفر» و «یک» استفاده می‌کند. به عنوان نتیجه‌ای از دگرشکلی، گاهی اوقات خطاها در حین ارسال ساخته می‌شوند. پیشامدهای زیر را در نظر بگیرید:

$u_۰$: یک «صفر» فرستاده شده باشد

یک «یک» فرستاده شده باشد u_1:
یک «صفر» دریافت شده باشد v_0:
یک «یک» دریافت شده باشد v_1:

احتمالهای زیر داده شده‌اند:

$$p(u_0)=\frac{۱}{۲}\ ,\qquad P(v_0|u_0)=\frac{۳}{۴}\ ,\qquad P(v_0|u_1)=\frac{۱}{۲}$$

(الف) وقتی شما می‌فهمید که کدام نماد دریافت شده در حالی که شما می‌دانید که «صفر» ارسال شده است آیا چه مقدار اطلاع دریافت می‌کنید؟

(ب) وقتی شما می‌فهمید که کدام نماد دریافت شده در حالی که شما می‌دانید که کدام نماد ارسال شده است آیا چه مقدار اطلاع کسب می‌کنید؟

(پ) وقتی کسی به شما بگوید که کدام نماد فرستاده شده و کدام نماد دریافت شده است، مقدار اطلاعی را که دریافت می‌کنید تعیین کنید.

(ت) وقتی به شما گفته شود که کدام نماد فرستاده شده در حالی‌که شما می‌دانید که کدام نماد دریافت شده است، مقدار اطلاع دریافتی را تعیین کنید.

۸.۱ جوابها

۱.۱ وقتی دو تاس را پرتاب می‌کنیم $۳۶=۶^۲$ برآمد ممکن هر یک با احتمال رخداد برابر یعنی $\frac{۱}{۳۶}$ وجود دارد. از هر پرتابی می‌توان مقدار اطلاعی برابر با

$$H(X)=\log ۳۶ = ۲٫۴۸ \text{ پرتاب / بیت}.$$

کسب نمود. چون می‌دانیم که مجموع وجوه برابر ۷ است، از ۳۶ امکان تعداد ۶تا باقی می‌ماند که هنوز می‌توانند رخ دهد. یعنی ۱-۶، ۶-۱، ۲-۵، ۵-۲، ۳-۴، ۴-۳. بنابراین هنوز عدم‌حتمیت باقی‌مانده وجود دارد که می‌توان مقدار اطلاعی برابر با

$$H'(X)=\log ۶ = ۱٫۲۴ \text{ پرتاب / بیت}.$$

دریافت کرد. چون فرض شده است که مجموع وجوه ۷ باشد مقدار اطلاعی که از این امر به دست می‌آید برابر است با

$$H(X)-H'(X)=\log ۳۶-\log ۶=\log ۶=۱٫۲۴ \text{ پرتاب / بیت}.$$

۲.۱ (الف) احتمال این‌که توپ سفید یا سیاه استخراج کنیم برابر است با

$$p(w)=\frac{n-m}{n}\ ,\qquad p(bl)=\frac{m}{n}.$$

بنابراین مقدار اطلاع دریافتی از آزمایش X برابر است با

$$\begin{aligned}H(\mathrm{X}) &= -p(w_{\mathrm{X}})\log p(w_{\mathrm{X}})-p(bl_{\mathrm{X}})\log p(bl_{\mathrm{X}})\\ &= -\frac{n-m}{n}\log\left[\frac{n-m}{n}\right]-\frac{m}{n}\log\frac{m}{n},\end{aligned}$$

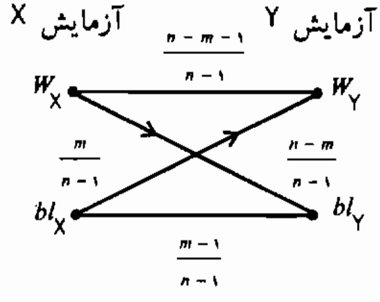
که در آن $p(w_{\mathrm{X}})$ و $p(bl_{\mathrm{X}})$ به ترتیب احتمالهای استخراج یک توپ سفید و یک توپ سیاه می‌باشند.

(ب) یک توپ به تصادف و بدون جای‌گذاری بیرون می‌آوریم، بنابراین هنوز $n-۱$ توپ برای آزمایش Y در گلدان باقی‌مانده است. اکنون بایستی بین امکان این که یک توپ سیاه و یا سفید در آزمایش X استخراج شده باشد تمایز قائل شد. یعنی، احتمالهای شرطی باید تعیین شوند. این احتمالها عبارتند از

$$p(w_{\mathrm{Y}}|\,w_{\mathrm{X}})=\frac{n-m-۱}{n-۱}\ ,\qquad p(bl_{\mathrm{Y}}|\,bl_{\mathrm{X}})=\frac{m-۱}{n-۱},$$

$$p(w_{\mathrm{Y}}|\,bl_{\mathrm{X}})=\frac{n-m}{n-۱}\ ,\qquad p(bl_{\mathrm{Y}}|\,w_{\mathrm{X}})=\frac{m}{n-۱}.$$

شکل (۱. ۱۰) را ببینید.



شکل ۱. ۱۰- احتمالهای شرطی نسبت به آزمایشهای X و Y تمرین (۲.۱)

رنگ توپ مربوط به آزمایش X معلوم نیست، بنابراین دو امکان w_{X} و bl_{X} وجود دارد. از این مطلب نتیجه می‌شود که

$$p(w_{\mathrm{Y}})=p(w_{\mathrm{X}})p(w_{\mathrm{Y}}|\,w_{\mathrm{X}})+p(bl_{\mathrm{X}})p(w_{\mathrm{Y}}|\,bl_{\mathrm{X}})$$

$$=\frac{n-m}{n}\frac{n-m-۱}{n-۱}+\frac{m}{n}\frac{n-m}{n-۱}=\frac{n-m}{n}=p(w_{X}).$$

به طور مشابه نتیجه می‌شود که $p(bl_{Y})=p(bl_{X})$. بنابراین برای مقدار اطلاع حاصل از این آزمایش نتیجه می‌شود که $H(Y)=H(X)$. این مطلب با توجّه به این که آزمایش X هیچ اطلاع واقعی که بتواند عدم‌حتمیت آزمایش Y را کاهش دهد نمی‌دهد، نیز دیده می‌شود.

(پ) می‌توان دو حالت تشخیص داد. اگر در آزمایش X یک توپ سفید استخراج شده باشد در این صورت مقدار اطلاع در آزمایش Y برابر است با

$$H(Y|w_{X})=-p(w_{Y}|w_{X})\log p(w_{Y}|w_{X})-p(bl_{Y}|w_{X})\log p(bl_{Y}|w_{X})$$

$$=-\frac{n-m-۱}{n-۱}\log\left[\frac{n-m-۱}{n-۱}\right]-\frac{m}{n-۱}\log\left[\frac{m}{n-۱}\right].$$

اگر توپ سیاه استخراج شده باشد، در این صورت داریم

$$H(Y|bl_{X})=-p(w_{Y}|bl_{X})\log p(w_{Y}|bl_{X})-p(bl_{Y}|bl_{X})\log p(bl_{Y}|bl_{X})$$

$$=-\frac{n-m}{n-۱}\log\left[\frac{n-m}{n-۱}\right]-\frac{m-۱}{n-۱}\log\left[\frac{m-۱}{n-۱}\right].$$

۳.۱ (الف) اگر رنگ قسمتی را که توپ در آن قرار می‌گیرد مشاهده کنیم، آزمایش می‌تواند سه برآمد ممکن داشته باشد؛ یعنی سبز، قرمز و سیاه با احتمالهای رخداد $p(\text{سبز})=\frac{۲}{۳۸}=\frac{۱}{۱۹}$، $p(\text{قرمز})=\frac{۱۸}{۳۸}=\frac{۹}{۱۹}$ و $p(\text{سیاه})=\frac{۱۸}{۳۸}=\frac{۹}{۱۹}$.

اگر تنها رنگ مورد توجّه باشد مقدار اطلاع برابر است با

$$H(\text{رنگ})=-\sum_{i}p_{i}\log p_{i}$$

$$=-\frac{۱}{۱۹}\log\frac{۱}{۱۹}-۲\frac{۹}{۱۹}\log\frac{۹}{۱۹}$$

$$=-\frac{۳۶}{۱۹}\log ۳+\log ۱۹=۱{,}۲۴\ \text{بیت}$$

(ب) مقدار اطلاع را با توجّه به این که هر بخش با شمارهٔ داده شده کاملاً تعیین می‌گردد می‌توان به دست آورد. تعداد ۳۸ بخش وجود دارد که هر کدام با احتمال برابر رخ می‌دهد، بنابراین

$$H(\text{شماره و رنگ}) = H(\text{شماره}) = \log 38 = 5{,}25 \text{ بیت}$$

مقدار اطلاع شرطی $H(\text{رنگ}|\text{شماره})$ به طور وضوح برابر صفر است، این مطلب واضح است زیرا برای یک شمارهٔ داده شده رنگ به طور خودکار معلوم است.

(پ) اطلاع شرطی اگر رنگ معلوم باشد برابر است با

$$H(\text{شماره}|\text{رنگ}) = H(\text{شماره و رنگ}) - H(\text{رنگ})$$

$$= 5{,}25 - 1{,}24 = 4{,}01 \text{ بیت.}$$

۴.۱ (الف) احتمالهای بیرون آوردن یک توپ سیاه و یا سفید برابر است با $p(bl_X) = \frac{1}{3}$ و $p(w_X) = \frac{2}{3}$.

وقتی توپ را به تصادف بیرون می‌آوریم مقدار اطلاع یا عدم‌حتمیت دریافتی برابر است با

$$H(X) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0{,}92 \text{ بیت.}$$

(ب) برای آزمایش Y، اگر برآمد X سیاه باشد احتمالها عبارتند از

$$p(bl_Y|bl_X) = \frac{4}{14} = \frac{2}{7} \quad \text{و} \quad p(w_Y|bl_X) = \frac{10}{14} = \frac{5}{7}$$

بنابراین داریم

$$H(Y|bl_X) = -\frac{2}{7}\log\frac{2}{7} - \frac{5}{7}\log\frac{5}{7} = 0{,}86 \text{ بیت.}$$

(پ) اگر برآمد X سفید باشد، در این صورت به طور مشابه نتیجه می‌شود که $p(bl_Y|w_X) = \frac{5}{14}$ و $p(w_Y|w_X) = \frac{9}{14}$ و

$$H(Y|w_X) = -\frac{9}{14}\log\frac{9}{14} - \frac{5}{14}\log\frac{5}{14} = 0{,}94 \text{ بیت.}$$

(ت) عدم‌حتمیت در آزمایش Y مجموع موزون نتایج (ب) و (پ) می‌باشد، یعنی

$$H(Y|X) = p(bl_X)H(Y|bl_X) + p(w_X)H(Y|w_X)$$

$$= \frac{1}{3}\,0{,}86 + \frac{2}{3}\,0{,}94 = 0{,}91 \text{ بیت.}$$

۵.۱ (الف) چهار وضعیت ممکن قبول شدن، رد شدن، ماشین داشتن و ماشین نداشتن را به ترتیب با s، $\bar{s}$، c و $\bar{c}$ نمایش می‌دهیم. وقتی یک نتیجهٔ امتحان اعلان می‌شود

یک مقدار اطلاع به دست می‌آید.

$$H(\text{نتیجه}) = -p(s)\log p(s) - p(\bar{s})\log p(\bar{s})$$

$$= -\frac{۳}{۴}\log\frac{۳}{۴} - \frac{۱}{۴}\log\frac{۱}{۴} = ۰/۸۱ \text{ بیت}.$$

(ب) اگر یک دانشجو که قبول شده است اعلان کند که ماشین دارد یا ندارد، در این صورت دو امکان c و $\bar{c}$ با احتمالهای معلوم وجود دارد. بنابراین

$$H(\text{ماشین دارد}|\text{قبول شده}) = -p(c|s)\log p(c|s) - p(\bar{c}|s)\log p(\bar{c}|s)$$

$$= -\frac{۱}{۱۰}\log\frac{۱}{۱۰} - \frac{۹}{۱۰}\log\frac{۹}{۱۰} = ۰/۴۷ \text{ بیت}.$$

(پ) در کل چهار امکان وجود دارد. احتمالهای نظیر عبارتند از

$$p(s,c) = \frac{۳}{۴}\times\frac{۱}{۱۰} = \frac{۳}{۴۰},$$

$$p(s,\bar{c}) = \frac{۳}{۴}\times\frac{۹}{۱۰} = \frac{۲۷}{۴۰},$$

$$p(\bar{s},c) = \frac{۱}{۴}\times\frac{۱}{۲} = \frac{۱}{۸},$$

$$p(\bar{s},\bar{c}) = \frac{۱}{۴}\times\frac{۱}{۲} = \frac{۱}{۸}.$$

مقدار اطلاعی که به واسطهٔ اعلان نتیجهٔ امتحان به دست می‌آید و امکان مالکیّت ماشین عبارت است از

$$H(\text{ماشین دارد , نتیجه}) = -\frac{۳}{۴۰}\log\frac{۳}{۴۰} - \frac{۲۷}{۴۰}\log\frac{۲۷}{۴۰} - ۲\times\frac{۱}{۸}\log\frac{۱}{۸}$$

$$= ۱/۴۱ \text{ بیت}.$$

عدم‌حتمیت باقی‌مانده دربارهٔ مالکیت ماشین، اگر نتیجهٔ امتحان، داده شده باشد برابر است با

$$H(\text{ماشین‌دار}|\text{نتیجه}) = H(\text{نتیجه و مالکیت ماشین}) - H(\text{نتیجه})$$

$$= ۱/۴۱ - ۰/۸۱ = ۰/۶۰ \text{ بیت}.$$

همچنین این نتیجه را مستقیماً با محاسبهٔ مقدار اطلاع شرطی برطبق تعریف مقدار اطلاع شرطی می‌توان به دست آورد.

$$H(\text{نتیجه}|\text{ماشین‌دار}) = \frac{۳}{۴}\left(-\frac{۱}{۱۰}\log\frac{۱}{۱۰} - \frac{۹}{۱۰}\log\frac{۹}{۱۰}\right) + \frac{۱}{۴}\left(-\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۲}\log\frac{۱}{۲}\right) = ۰٫۶۰ \text{ بیت}.$$

۶.۱ (الف) اگر او بور باشد برای رنگ چشم دو امکان، یعنی آبی و غیرآبی به ترتیب با احتمال $\frac{۱}{۴}$ و $\frac{۳}{۴}$ وجود دارد. بنابراین مقدار اطلاع شرطی دریافتی به این شرط که او بور است برابر است با

$$H(\text{رنگ چشم}|\text{بور}) = -\frac{۳}{۴}\log\frac{۳}{۴} - \frac{۱}{۴}\log\frac{۱}{۴} = ۰٫۸۱ \text{ بیت}.$$

(ب) برای این که قادر باشیم به این سؤال پاسخ دهیم باید احتمالهای $p(\text{بور}|\text{آبی})$ و $p(\text{غیربور}|\text{آبی})$ را تعیین کنیم. این احتمالها را با کمک فرمول بیز می‌توان به دست آورد؛ داریم

$$p(\text{بور}|\text{آبی}) = \frac{p(\text{بور})\ p(\text{آبی}|\text{بور})}{p(\text{آبی})} = \frac{\frac{۱}{۴}\times\frac{۳}{۴}}{\frac{۱}{۲}} = \frac{۳}{۸}$$

چون یک دختر بور است یا بور نیست، داریم

$$p(\text{بور}|\text{آبی}) + p(\text{غیربور}|\text{آبی}) = ۱$$

که نتیجه می‌شود

$$p(\text{غیربور}|\text{آبی}) = \frac{۵}{۸}$$

در این صورت مقدار اطلاع شرطی که دریافت می‌شود برابر است با

$$p(\text{رنگ مو}|\text{آبی}) = -\frac{۳}{۸}\log\frac{۳}{۸} - \frac{۵}{۸}\log\frac{۵}{۸} = ۰٫۹۵ \text{ بیت}.$$

(پ) اگر رنگ مو و همچنین رنگ چشمش معلوم باشد، می‌توان از پیشامد توأم با چهار برآمد ممکن صحبت کرد. با بهره‌وری از نتایج بخش-فرعی قبل احتمالهای این برآمدها را پیدا می‌کنیم که عبارتند از: $p(\text{آبی و بور}) = \frac{۳}{۱۶}$، $p(\text{غیرآبی و بور}) = \frac{۱}{۱۶}$، $p(\text{آبی و غیربور}) = \frac{۵}{۱۶}$ و $p(\text{غیرآبی و غیربور}) = \frac{۷}{۱۶}$.

مقدار اطلاع دریافتی برابر است با

$H($رنگ مو و رنگ چشم$) = -\frac{۳}{۱۶}\log\frac{۳}{۱۶} - \frac{۱}{۱۶}\log\frac{۱}{۱۶} - \frac{۵}{۱۶}\log\frac{۵}{۱۶} - \frac{۷}{۱۶}\log\frac{۷}{۱۶} = ۱٫۷۵$ بیت.

۷.۱ (الف) چهار وضعیت ممکن، یعنی ترکیب مناسب بودن یا نبودن و رد شدن و نشدن را می‌توان به عنوان نقطهٔ شروع در نظر گرفت. اگر این وضعیتها را با s، $\bar{s}$، r و $\bar{r}$ نشان دهیم، در این صورت داریم

$$p(\bar{s}) = \frac{۱}{۴}\,, \quad p(r|\bar{s}) = \frac{۳}{۴}\,, \quad p(r) = \frac{۱}{۲}.$$

از این رو داریم

$$p(s) = ۱ - p(\bar{s}) = \frac{۳}{۴},$$

$$p(\bar{r}|\bar{s}) = ۱ - \frac{۳}{۴} = \frac{۱}{۴},$$

$$p(\bar{r}) = ۱ - p(r) = \frac{۱}{۲}.$$

به علاوه بنابر فرمول بیز داریم

$$p(\bar{s}|r) = \frac{p(r|\bar{s})p(\bar{s})}{p(r)} = \frac{\frac{۳}{۴}\times\frac{۱}{۴}}{\frac{۱}{۲}} = \frac{۳}{۸}.$$

چون $p(\bar{s}|r) + p(s|r) = ۱$، نتیجه می‌شود که

$$p(s|r) = ۱ - \frac{۳}{۸} = \frac{۵}{۸}.$$

به طور مشابه می‌توان محاسبه کرد که

$$p(\bar{s}|\bar{r}) = \frac{p(\bar{r}|\bar{s})p(\bar{s})}{p(\bar{r})} = \frac{\frac{۱}{۴}\times\frac{۱}{۴}}{\frac{۱}{۲}} = \frac{۱}{۸},$$

$$p(s|\bar{r}) = ۱ - p(\bar{s}|\bar{r}) = \frac{۷}{۸}.$$

سرانجام، احتمالهای ترکیبی هر یک از چهار ترکیب را می‌توان از رابطهٔ کلّی زیر به دست آورد

$$r_{ij} = p_i q_{ji} = q_j p_{ij}\,.$$

از این رو داریم

$$p(s,r)=p(r)p(s|r)=\frac{۱}{۲}\times\frac{۵}{۸}=\frac{۵}{۱۶},$$
$$p(\bar{s},r)=p(r)p(\bar{s}|r)=\frac{۱}{۲}\times\frac{۳}{۸}=\frac{۳}{۱۶},$$
$$p(s,\bar{r})=p(s)p(\bar{r}|s)=\frac{۳}{۴}\times\frac{۷}{۱۲}=\frac{۷}{۱۶},$$
$$p(\bar{s},\bar{r})=p(\bar{s})p(\bar{r}|\bar{s})=\frac{۱}{۴}\times\frac{۱}{۴}=\frac{۱}{۱۶}.$$

نتایج را می‌توان در یک نمودار نشان داد، که در آن احتمالها در ۱۶ ضرب شده‌اند (شکل (۱۱.۱) را ببینید).

چون تنها نتیجهٔ گزینش بدون هیچ‌گونه ویژگی بیشتری نام برده شده است، دو امکان برای نتایج گزینش وجود دارد. مقدار اطلاع برابر است با

$$H(\text{گزینش}\,|\,\text{نامناسب})=-p(r|\bar{s})\log p(r|\bar{s})-p(\bar{r}|\bar{s})\log p(\bar{r}|\bar{s})$$
$$=-\frac{۳}{۴}\log\frac{۳}{۴}-\frac{۱}{۴}\log\frac{۱}{۴}=۰٫۸۱\ \text{بیت}.$$

(ب) اگر گزینش با پرتاب یک سکه انجام شود، هر دانشجو ۵۰٪ شانس رد شدن خواهد داشت. همچنین به عنوان نتیجهٔ تمام احتمالهای $p(r|s)$، $p(r|\bar{s})$، $p(\bar{r}|s)$ و $p(\bar{r}|\bar{s})$ بدون توجّه به شرط s یا $\bar{s}$ برابر $\frac{۱}{۲}$ می‌شود؛ مقدار اطلاع برابر می‌شود با:

$$H(\text{گزینش}\,|\,\text{نامناسب})=H(\text{گزینش})=-\frac{۱}{۲}\log\frac{۱}{۲}-\frac{۱}{۲}\log\frac{۱}{۲}$$
$$=۱\ \text{بیت}.$$

(پ) چون مناسب بودن یا نبودن، نقشی در (ب) بازی نمی‌کند، دانشجو نمی‌تواند از معلوماتش استفاده کند؛ یعنی این که او نامناسب است. مقدار اطلاعی که در (ب) دریافت می‌کند برابر با اطلاعی است که بعد از پرتاب یک سکّه دریافت می‌کند و بنابراین برابر یک است. عدم‌حتمیت برای (الف) کوچکتر است چنان‌که او نیز اطلاع کمتری دریافت می‌کند.

	r	$\bar{r}$	
s	۵	۷	۱۲
$\bar{s}$	۳	۱	۴
	۸	۸	

شکل ۱۱.۱- احتمالهای توأم تمرین (۷.۱)

۸.۱ (الف) با استفاده از ماتریس داده شده مستقیماً نتیجه می‌شود که

$$H(X,Y) = -\sum_{i=1}^{3}\sum_{j=1}^{3} r_{ij}\log r_{ij}$$
$$= -(2\times\frac{7}{24}\log\frac{7}{24} + 4\times\frac{1}{24}\log\frac{1}{24} + \frac{1}{4}\log\frac{1}{4}) = 2{,}30 \text{ بیت}.$$

(ب) چون برای تمام jها داریم

$$q_j = \sum_{i=1}^{3} r_{ij}$$

نتیجه می‌شود که $q(y_1) = q(y_2) = q(y_3) = \frac{1}{3}$. در این صورت مقدار اطلاع برابر می‌شود با

$$H(Y) = \log 3 = 1{,}58 \text{ بیت}.$$

(پ) می‌خواهیم $H(X|Y)$ را محاسبه کنیم. این به سادگی از رابطهٔ $H(X,Y) = H(Y) + H(X|Y)$ محاسبه می‌شود و بیت $H(X|Y) = 0{,}72$ به دست می‌آید. همچنین می‌توان احتمالهای شرطی p_{ij} را از r_{ij} و q_j به دست آورد و در رابطهٔ زیر قرار داد

$$H(X|Y) = -\sum_{i=1}^{3}\sum_{j=1}^{3} r_{ij}\log(p_{ij}).$$

۹.۱ (الف) داریم: $p(v_1|u_0) = 1 - p(v_0|u_0) = \frac{1}{4}$.

از این رو برای عدم‌حتمیت با توجّه به نماد دریافت شده به فرض این که «صفر» ارسال شده باشد، به دست می‌آوریم

$$H(V|u_0) = -p(v_0|u_0)\log p(v_0|u_0) - p(v_1|u_0)\log p(v_1|u_0)$$
$$= -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0{,}82 \text{ بیت}.$$

(ب) ابتدا احتمالهای توأم را حساب می‌کنیم؛ داریم: $p(u_0,v_0) = p(v_0|u_0)p(u_0) = \frac{3}{8}$.
به طور مشابه داریم

$$p(u_0,v_1) = \frac{1}{8} \quad ، \quad p(u_1,v_0) = \frac{1}{4} \quad ، \quad p(u_1,v_1) = \frac{1}{4}$$

اکنون مقدار اطلاع با توجّه به نماد دریافتی به شرط نماد ارسال شده به صورت زیر

می‌باشد

$$H(V|U) = -\sum_{i=0}^{1}\sum_{j=0}^{1} p(u_i, v_j)\log p(v_j|u_i)$$

$$= -\frac{3}{8}\log\frac{3}{4} - \frac{1}{8}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{2} = 0{,}91 \text{ بیت.}$$

(پ) **روش I:** با قرار دادن احتمالهای توأم در فرمول برای اطلاع توأم نتیجه می‌شود:

$$H(U,V) = -\sum_{i=0}^{1}\sum_{j=0}^{1} p(u_i, v_j)\log p(u_i, v_j) = 1{,}91 \text{ بیت.}$$

روش II: چون $p(u_0) = p(u_1) = \frac{1}{2}$، مقدار $H(U)$ با توجّه به نماد ارسال شده برابر است با: $H(U) = 1$ بیت. اکنون نتیجه می‌شود که

$$H(U,V) = H(U) + H(V|U) = 1 + 0{,}91 = 1{,}91 \text{ بیت.}$$

در این حالت، این روش سریعتر از روش I می‌باشد.

(ت) چون می‌توان به دست آورد که $p(v_0) = \frac{5}{8}$ و $p(v_1) = \frac{3}{8}$، برای اطلاع $H(V)$ نسبت به نماد دریافت شده نتیجه می‌شود که

$$H(V) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{3}{8} = 0{,}96 \text{ بیت.}$$

از این رو داریم

$$H(U|V) = H(U,V) - H(V) = 1{,}91 - 0{,}96 = 0{,}95 \text{ بیت.}$$

۲

منبع اطلاع گسستهٔ بی‌حافظه

۱.۲ منبع اطلاع گسسته

منبع اطلاع گسسته منبعی است که دنباله‌ای از نمادها را (که گاهی اوقات حروف نیز نامیده می‌شوند) تولید می‌کند، که در آن هر نماد متعلق به همان مجموعه نمادهای ممکن است. این مجموعه نمادهای ممکن را ***الفبای منبع*** می‌نامند. نمادها را با u_1، u_2، ... ، u_n و الفبا را با

$$U = \{u_1, \ldots, u_i, \ldots, u_n\}$$

نمایش می‌دهیم. این نمادها در نقاط زمانی گسسته تولید می‌شوند. به این دلیل همراه با این حقیقت که الفبای منبع متناهی است آن را منبع اطلاع گسسته می‌نامیم. یک گروه متوالی از نمادها را ***پیام*** یا ***واژه*** می‌نامیم.

برخی تشابهات با عبارات نوشتنی وجود دارد. در این صورت منبع اطلاع U را می‌توان به صورت الفبا با ۲۶ حرفش، فاصله و احتمالاً چند علامت نشان‌گذاری در نظر گرفت. در عبارت نوشتنی واژه‌ها شامل گروهی از حروفند که با یک فاصله از هم جدا می‌شوند.

واژه‌ها یا پیامهایی به طول l را با v نمایش می‌دهیم. چون الفبا شامل n نماد است، تعداد پیامهای ممکن برابر n^l است. مجموعهٔ $V = \{v_1, \ldots, v_j, \ldots, v_{n^l}\}$ مجموعهٔ تمام پیامهای ممکن می‌باشد.

با تمایز بین نمادها و پیامها می‌توان منبع اطلاع را به دو طریق مورد بررسی قرار داد: در سطح نمادی و در سطح پیام. فرض کنید که منبع اطلاع تصادفی است، یعنی این که

نمادهـای الفبـای U هـر یـک بـا احتمـال معیّنـی رخ می‌دهـد. ایـن احتمالهـا را بـا $p(u_1)=p_1$ ، $p(u_2)=p_2$ و الی آخر نشان می‌دهیم، با توزیع احتمال مربوط به U مواجهیم:

$$P=\{p_1,\ldots,p_i,\ldots,p_n\}$$

در سرتاسر این کتاب صراحتاً فرض می‌کنیم که احتمالهای مورد ملاحظه با گذشـت زمان بدون تغییر باقی می‌مانند. در کاربردهای بسیاری این فرض تأیید شده اسـت. در ایـن صورت می‌گوییم که نمادها از یک دنبالهٔ تصادفی تولید شده‌اند که ***مانا*** است.

جنبهٔ دوم که مهم می‌باشد وابستگی متقابل نمادهای متوالی در یک پیام است. در این صورت از حافظهٔ منبع سخن می‌گوییم. در این فصل یک ***منبع بی‌حافظه*** را بررسی خواهیـم کرد؛ یعنی این که بگوییم نمادهای تولید شده به طور آماری مستقلّند.

با ملاحظهٔ منبع اطلاع در سطح نمادی، مقدار اطلاعی که توسط منبع گسسته بی‌حافظه تولید می‌شود برابر است با

$$H(U)=-\sum_{i=1}^{n}p_i\log p_i \text{ بیت / نماد}. \qquad (1.2)$$

ماکسیمم مقدار اطلاعی که می‌توان توسط یک منبع گسستهٔ بی‌حافظـه تولیـد کـرد عبارت است از

$$\max_u H(U)=\log n \text{ بیت / نماد}. \qquad (2.2)$$

همان‌طور که در قضیهٔ (۱.۱) نشان دادیم اگر احتمال وقوع همهٔ نمادها برابر باشـند، یعنی اگر برای همهٔ iها داشته باشیم $p_i=\frac{1}{n}$ این ماکسیمم به دست می‌آید. با مقایسهٔ مقدار اطلاع $H(U)$ با ماکسیمم ممکن اثری از حشو منبع به دست می‌آوریم

تعریف ۱.۲

حشو یک منبع اطلاع گسستهٔ بی‌حافظه را به صورت زیرتعریف می‌کنیم

$$red=1-\frac{H(U)}{\max\limits_u H(U)}=1-\frac{H(U)}{\log n}, \qquad (3.2)$$

که در آن $H(U)$ مقدار اطلاع (حجم اطلاع) یک منبع اطلاع با الفبای منبع باحجم n می‌باشد. ●

واضح است اگر منبعی نمادهایی با احتمال وقوع برابر تولیـد کنـد در ایـن صـورت داریم $H(U)=\max H(U)$، به قسمی که برای حشو به دست می‌آوریم $red=0$. در حـالتی که در آن منبع تنها یک نماد تولید می‌کند، یعنی یک نماد دارای احتمال یـک اسـت در

حالی که بقیه احتمال صفر دارند، در این صورت $H(U)=0$ و در نتیجه داریم: $red=1$.
بنابراین مقدار حشو بین ۰ و ۱ تغییر خواهد کرد.

منبعی که می‌تواند دو نماد تولید کند **منبع دودویی** نامیده می‌شود. از مطالب قبلی نتیجه می‌شود که چنین منبعی می‌تواند ماکسیمم $\log 2 = 1$ بیت / نماد تولید کند، لکن برای حالت کلّی، احتمالهای نماد به ترتیب p و $(1-p)$ هستند و مقدار اطلاع تولید شده کمتر خواهد بود، یعنی

$$H(U) = -p\log p - (1-p)\log(1-p) \text{ بیت / نماد},$$

و از این رو حشو بزرگتر خواهد بود.

مثال ۱.۲

یک منبع دودویی نمادهای ۰ و ۱ را با احتمالهای $\frac{1}{4}$ و $\frac{3}{4}$ تولید می‌کند. بنابراین الفبا عبارت است از $U=\{0,1\}$ و علاوه بر این داریم

$$H(U) = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4} = 0{,}81 \text{ بیت / نماد}$$

و

$$\max_u H(U) = \log 2 = 1 \text{ بیت / نماد}.$$

در این صورت حشو برابر است با

$$red = 1 - \frac{H(U)}{\max\limits_u H(U)} = 1 - \frac{0{,}81}{1} = 0{,}19. \qquad \blacktriangle$$

اکنون تعیین خواهیم کرد که اگر پیامها را به جای نمادهای تکی در نظر بگیریم برای مقدار اطلاع چه نتیجه می‌شود. نخست مثال زیر را بررسی خواهیم کرد

مثال ۲.۲

منبع دودویی مثال (۱.۲) را در نظر بگیرید و فرض کنید که پیامهایی به طول ۳ تشکیل شده‌اند. اکنون همچنان که از شکل (۱.۲) پیداست ۸ پیام ممکن وجود دارد که با v_1، ...، v_8 نشان داده شده‌اند.

چون منبع بی‌حافظه است، نمادها به طور آماری مستقل می‌باشند و برای مثال پیام v_2 را بررسی می‌کنیم

$$p(001) = p(0)p(0)p(1) = \frac{1}{4}\frac{1}{4}\frac{3}{4} = \frac{3}{64}.$$

به عبارت دیگر، احتمال هر پیام با حاصل‌ضرب احتمالهای نمادهای تکی برابر است. احتمالهای همهٔ پیام در شکل (۱.۲) نشان داده شده‌اند. اکنون اگر مقدار اطلاع را نسبت بـه پیام حساب کنیم خواهیم داشت

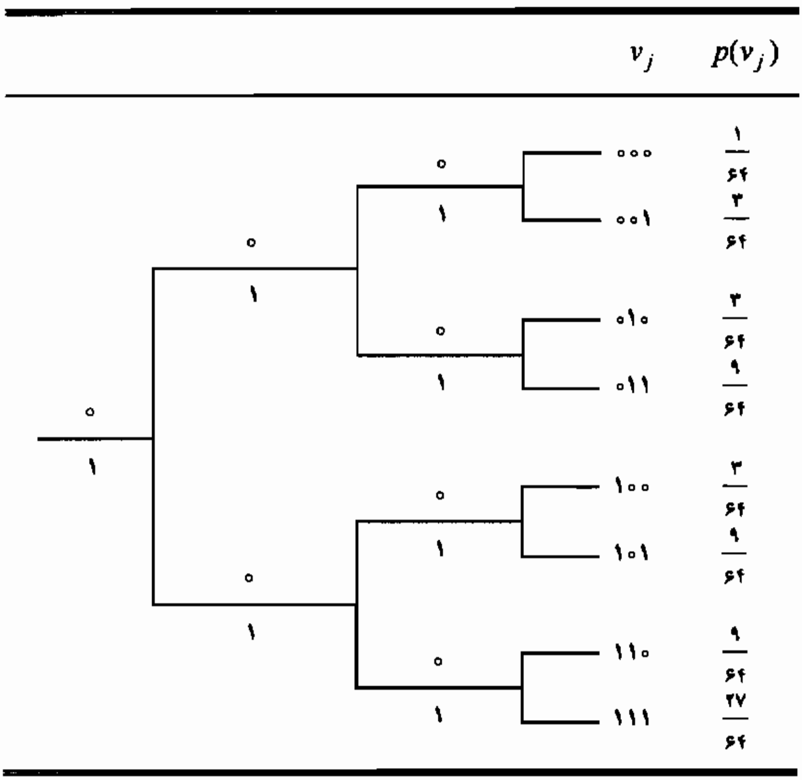
$$H(V) = -\frac{۱}{۶۴}\log\frac{۱}{۶۴} - ۳\times\frac{۳}{۶۴}\log\frac{۳}{۶۴} - ۳\times\frac{۹}{۶۴}\log\frac{۹}{۶۴} - \frac{۲۷}{۶۴}\log\frac{۲۷}{۶۴}$$

$$= ۲٫۴۵ \text{ پیام / بیت}.$$

توجّه کنید که

$$H(V) = ۳H(U).$$

▲



شکل ۱.۲- تعداد پیامهای ممکن به طول ۳ در حالت منبع دودویی

به طور کلّی می‌توان n^l پیام متفاوت به طول l از الفبای منبع U بـا حجم n فراهـم کرد. چون با منبع بی‌حافظه سروکار داریم، احتمال رخ‌داد هر پیام بـا حـاصل‌ضرب احتمالهای l نماد تکی که پیام را می‌سازند برابر است. مقدار اطلاع در سطح پیام به صورت

زیر است

$$H(V) = -\sum_{j=1}^{n^l} p(v_j)\log p(v_j). \qquad (۴.۲)$$

با نوشتن احتمالهای $p(v_j)$ برحسب $p(u_1)$، ...، $p(u_n)$ می‌توان به سادگی ثابت کرد که

$$H(V) = l\,H(U). \qquad (۵.۲)$$

بنابراین یک پیام به طول l شامل اطلاعی به اندازهٔ l برابر پیامی با طول ۱ می‌باشد. توجّه کنید علّت آن این است که نمادهای متوالی مستقلّند. بعداً به حالتی برمی‌گردیم که این مطلب برای آن درست نخواهد بود.

همچنین مقدار اطلاع یک منبع اطلاع را می‌توان برحسب بیت بر واحد زمان مثلاً ثانیه / بیت بیان کرد. در این صورت از *تولید* منبع سخن می‌گوییم. اگر همهٔ نمادها دارای زمان یکسان مثلاً t ثانیه باشند در این صورت تولید $H_t(U)$ منبع برابر است با

$$H_t(U) = \frac{1}{t}H(U) \text{ ثانیه / بیت}. \qquad (۶.۲)$$

اگر نمادها دارای مدّت یکسان نباشند مثلاً مانند حالت کد مُرس که در آن «خط فاصله» (-) زمان بیشتری از «نقطه» می‌گیرد مقدار متوسط زمان t به کار برده می‌شود.

۲.۲ کدگذاری منبع

یک منبع اطلاع بی‌حافظه را که از نمادهای الفبای منبع $U = \{u_1, u_2, \dots, u_n\}$ پیامها را می‌سازد در نظر می‌گیریم. به طور کلّی برطبق الگوی ارتباطی شکل (۹.۱) نخست *کاهش داده‌ها* را برای حذف اطلاعاتی که برای مقصد مناسب نیستند به کار می‌بریم. در این جا فرض می‌کنیم که منبع اطلاع پیامهایی که بیشتر از این نمی‌توان تقلیل داد تولید می‌کند.

به دلیل کارایی با حذف حشو موجود در پیام می‌خواهیم پیامها را در حدّ امکان به طور متراکم بیان کنیم. این فرایند به عنوان *کدگذاری منبع* شناخته می‌شود. تنها به چند مثال از کدهایی که می‌توان در کدگذاری پیامها به کار برد اکتفا می‌کنیم. چون فرض شده که منبع اطلاع بی‌حافظه باشد، کافی است که کدگذاری نمادهای مجزّای منبع را به جای پیامها در نظر بگیریم.

فرض کنید الفبای کد با $S = \{s_1, s_2, \dots, s_r\}$ داده شده باشد. اکنون کدهایی را جست‌وجو

می‌کنیم که ترکیب معیّنی از کدنمادها را برای هر نماد خروجی منبع، که یک ***کدواژه*** نامیده می‌شود، می‌دهد.

اگر همهٔ کدواژه‌ها متفاوت باشند آن را ***کد ناویژه*** می‌نامیم. اگر برای یک ردیف کدواژه نتیجه نیز ناویژه باقی بماند دارای کدی است که ***به طور یکتا قابل کدگشایی*** است. در حالت قابلیّت کدگشایی یکتا پیام دریافت شده بایستی امکان تفسیر یکتای تکی داشته باشد. سرانجام، در حالتی که یک کد به طور یکتا قابل کدگشایی است و هر نماد پیام را نیز بتوان مستقیماً، یعنی بدون نگاه کردن به کدنماد بعدی، کدگشایی کرد، در این صورت آن را یک ***کد فوری*** می‌نامیم.

مثال ۳.۲

یک منبع اطلاع دارای یک الفبا با چهار نماد منبع خروجی u_1، u_2، u_3 و u_4 می‌باشد. الفبای کد شامل دو نماد ۰ و ۱ است. کدواژه‌ها با استفاده از چهار سیستم کدگذاری متفاوت برطبق جدول زیر ساخته شده‌اند

	A	B	C	D
u_1	۰	۰۰	۰	۰
u_2	۱۱	۰۱	۱۰	۰۱
u_3	۰۰	۱۰	۱۱۰	۰۱۱
u_4	۰۱	۱۱	۱۱۱۰	۰۱۱۱

هر چهار کد ناویژه‌اند. کد A به طور یکتا قابل کدگشایی نیست. کدهای B، C و D به طور یکتا قابل کدگشایی‌اند: B به علّت این که همهٔ کدواژه‌ها طول مساوی دارند بنابراین فقط لازم است که دنبالهٔ متوالی از کدواژه‌ها را به گروه دو نمادی تقسیم کنیم، C بدین دلیل که هر کدواژه با یک ۰ ختم می‌شود که در این صورت به عنوان یک کاما عمل می‌کنند ("کد کاما")، D به این دلیل که هر کدواژه با یک ۰ شروع می‌شود. کد D قابل کدگشایی فوری نیست؛ زیرا همواره بایستی برای اولّین نماد کدواژهٔ بعدی صبر کرد قبل از این که بتوان کدواژهٔ جاری را کدگشایی کرد. ▲

شرط لازم و کافی برای یک کد فوری آن است که یک کدواژهٔ کامل هرگز شروع کدواژهٔ دیگری نیست. در این حالت از یک ***کد پیشوند*** نیز می‌توان نام برد.

از محدودیتهای گوناگون مشخص شده در بالا نتیجه می‌شود که یک رابطهٔ یک به یک

بین نمادهای منبع خروجی با الفبای $U=\{u_1,u_2,\dots,u_n\}$ و کدواژه‌ها وجـود دارد. بنـابراین برای سادگی، این کدواژه‌ها را نیز با u_1، u_2، ...، u_n نشان خواهیم داد. طول این کدواژه‌هـا را با l_1، l_2، ...، l_n نشان خواهیم داد. این طولها با تعداد کدنمادهایی که کدواژه را می‌سازند تعیین می‌گردند.

در قضیهٔ زیر بررسی می‌کنیم که یک کد چه شرطی باید داشته باشد تـا فـوراً قـابل کدگشایی باشد.

قضیهٔ ۱.۲ (نابرابری کرافت[1])

شرط لازم و کافی برای وجود یک کد فوری آن است که

$$\sum_{i=1}^{n} r^{-l_i} \le 1, \qquad (7.2)$$

که در آن r اندازهٔ الفبای کد و l_i، $i=1,\dots,n$، طول کدواژهٔ u_i می‌باشد.

برهان

فرض کنید تعداد کدواژه‌ها با طول ۱ برابر w_1 باشد. این تعداد حداکثر برابر r خواهد بود ($w_1 \le r$). کدواژه‌هایی که استفاده شده‌اند ممکن نیست شروع کدواژهٔ دیگری باشند. از این رو $r-w_1$ کدنماد اول باقی می‌ماند. از این رو برای تعداد کدواژه‌های با طول ۲ داریم

$$w_2 \le (r-w_1)r = r^2 - w_1 r.$$

به طور مشابه $w_3 \le \{(r-w_1)r - w_2\}r = r^3 - w_1 r^2 - w_2 r$.

و اگر m ماکسیمم طول کدواژه‌ها باشد، در این صورت داریم

$$w_m \le r^m - w_1 r^{m-1} - w_2 r^{m-2} - \cdots - w_{m-1} r.$$

با تقسیم آن بر r^m نتیجه می‌شود

$$0 \le 1 - w_1 r^{-1} - w_2 r^{-2} - \cdots - w_{m-1} r^{-m+1} - w_m r^{-m}.$$

یا

$$\sum_{j=1}^{m} w_j r^{-j} \le 1.$$

یعنی

$$\underbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}_{w_1}+\underbrace{\frac{1}{r^2}+\frac{1}{r^2}+\cdots+\frac{1}{r^2}}_{w_2}+\cdots+\underbrace{\frac{1}{r^m}+\frac{1}{r^m}+\cdots+\frac{1}{r^m}}_{w_m}\le 1$$

---
1. Kraft

ولی $w_1 + w_2 + \cdots + w_m = n$، یعنی کلّ تعداد کدواژه‌ها در نتیجه این نابرابر با نابرابری زیر یکی است

$$\sum_{i=1}^{n} r^{-l_i} \le 1 \qquad \blacksquare$$

توجّه کنید نابرابری کرافت نشان می‌دهد که یک کد فوری وجود دارد که دارای کدواژه‌ای به طول l_i می‌باشد. با وجود این، معنای آن این نیست که هر کدی که در نابرابری صدق کند یک کد فوری است.

مرحلهٔ بعد مربوط به روشی می‌شود که در آن انتخاب خاصّی برای طول l_i در نظر گرفته می‌شود. ممکن است مایل باشیم طول کدواژه‌ها بستگی به احتمال رخ‌داد پیام داشته باشد برای این که استفادهٔ بهینه از کانال را تضمین کند. یعنی، ترجیحاً به پیامهای با احتمال رخ‌داد زیاد کدواژهٔ کوتاهتری از پیامهای با احتمال رخ‌داد کم نسبت خواهیم داد.

مثال ۴.۲

با کد مُرس که در آن حروف الفبا به کدواژه‌هایی مرکّب از نقطه و خط برگردانده

نماد	احتمال	کد مُرس	نماد	احتمال	کد مُرس
A	۰٫۰۶۴۲	·–	N	۰٫۰۵۷۴	–·
B	۰٫۰۱۲۷	–··	O	۰٫۰۶۳۲	–––
C	۰٫۰۲۱۸	–·–·	P	۰٫۰۱۵۲	·––·
D	۰٫۰۳۱۷	–··	Q	۰٫۰۰۰۸	––·–
E	۰٫۱۰۳۱	·	R	۰٫۰۴۸۴	·–·
F	۰٫۰۲۰۸	··–·	S	۰٫۰۵۱۵	···
G	۰٫۰۱۵۲	––·	T	۰٫۰۷۹۶	–
H	۰٫۰۴۶۷	····	U	۰٫۰۲۲۸	··–
I	۰٫۰۵۷۵	··	V	۰٫۰۰۸۳	···–
J	۰٫۰۰۰۸	·––	W	۰٫۰۱۷۵	·––
K	۰٫۰۰۴۹	–·––	X	۰٫۰۰۱۴	–··–
L	۰٫۰۳۲۱	·–··	Y	۰٫۰۱۶۴	–·––
M	۰٫۰۱۹۸	––	Z	۰٫۰۰۰۵	––··
			Space	۰٫۱۸۵۹	

شکل ۲.۲- کد مُرس (احتمالها برای زبان انگلیسی می‌باشند)

می‌شود، کدواژه برای حروفی که مکرراً ظاهر می‌شوند (برای مثال مانند حرف e) طوری انتخاب شده‌اند که در حدّ ممکن شامل چند نقطه و خط از یک طرف و از طرف دیگر ترجیحاً شامل نقاط باشند زیرا نقطه زمان کوتاهتری از خط دارد. یک نقطه دو واحد زمان صرف می‌کند، در حالی که یک خط به ۴ واحد زمان نیاز دارد. فاصلهٔ حرفی شامل ۳ واحد زمان است. در شکل (۲.۲) کد مُرس ارائه شده است. ▲

قضیهٔ زیر رابطهٔ بین متوسط طول کدواژهٔ L و مقدار اطلاع یک منبع اطلاع را بیان می‌کند.

قضیهٔ ۲.۲ (قضیهٔ کدگذاری منبع)

مجموعهٔ n کدواژه u_i با توزیع احتمال $P=(p_1,\dots,p_n)$، $p_i>0$ را برای تمام iها، که در آن تمام کدواژه‌ها ترکیبی از نمادهای الفبای کد $S=(s_1,s_2,\dots,s_r)$ می‌باشند در نظر بگیرید. اگر نابرابری کرافت برقرار باشد آن‌گاه

$$\frac{H(U)}{\log r}\le L, \tag{۸.۲}$$

که در آن L متوسط طول کدواژه است و به صورت زیر تعریف می‌شود:

$$L=\sum_{i=1}^{n}p_i l_i\,, \tag{۹.۲}$$

و l_i طول کدواژهٔ u_i می‌باشد.

برابری برقرار می‌شود اگر و تنها اگر برای $i=1,\dots,n$، $p_i=r^{-l_i}$.

برهان

داریم

$$H(U)-L\log r=-\sum_{i=1}^{n}[p_i\log p_i+p_i l_i\log r]$$

$$=\sum_{i=1}^{n}p_i\log\left\{\frac{1}{p_i r^{l_i}}\right\}=\sum_{i=1}^{n}p_i\,\frac{\ln\left\{\frac{1}{p_i r^{l_i}}\right\}}{\ln 2}. \tag{۱۰.۲}$$

چون برای $a>0$ داریم

$$\ln a\le a-1$$

اگر $a=1$ برابری برقرار می‌شود (شکل (۲.۱) را ببینید)، از معادلهٔ (۱۰.۲) نتیجه می‌شود که

$$\sum_{i=1}^{n} p_i \ln\left(\frac{1}{p_i r^{l_i}}\right) \leq \sum_{i=1}^{n} p_i\left(\frac{1}{p_i r^{l_i}}-1\right) = \sum_{i=1}^{n} r^{-l_i} - 1.$$

چون نابرابری کرافت برقرار می‌باشد بنابراین نتیجه می‌شود که

$$H(U) - L\log r \leq 0,$$

که دقیقاً فرمول (۸.۲) را می‌دهد. شرط برابری را می‌توان مستقیماً از این به دست آورد. ■

آنچه که قضیهٔ (۲.۲) دقیقاً بیان می‌کند این است که متوسّط طول هرگز نمی‌تواند کوچکتر از مقدار اطلاع (در مبنای r) یک منبع اطلاع باشد. وقتی طول l_iها طوری انتخاب شده باشند که علامت برابری در معادلهٔ (۸.۲) برقرار باشد کوچکترین طول متوسط کدواژه‌ها به دست می‌آید.

این مطلب درست است اگر برای همهٔ iها

$$p_i = r^{-l_i} \qquad \text{یا} \qquad l_i = -\log_r p_i. \tag{۱۱.۲}$$

این را تنها اگر $-\log_r p_i$ عدد درست باشد می‌توان به دست آورد، چون l_i طول کدواژه‌ای است که از تعداد صحیحی از نمادها هریک با طول یک ساخته شده است. اگر $-\log_r p_i$ عدد صحیحی نباشد کدگذاری بهینه امکان‌پذیر نخواهد بود. بنابراین واضح به نظر می‌رسد که عدد صحیحی که بلافاصله بالای این طول l_i قرار دارد انتخاب کنیم؛ از این رو l_i را به طریق زیر انتخاب می‌کنیم

$$-\log_r p_i \leq l_i < -\log_r p_i + 1. \tag{۱۲.۲}$$

توجّه کنید که در حالت یک کد بهینه، یعنی $l_i = -\log_r p_i$، داریم $\sum_i r^{-l_i} = 1$ (با نابرابری کرافت مقایسه کنید) و در حالتهای دیگر داریم $\sum_i r^{-l_i} < 1$.

از نقطه نظر قضیهٔ (۲.۲)، اکنون ساختن معیاری برای کیفیت یک کد نیز کار ساده‌ای است. با نزدیک شدن بیشتر $\dfrac{H(U)}{L\log r}$ به مقدار ۱ (یعنی علامت برابری در معادلهٔ (۸.۲)) کد *کاراتری* داریم.

تعریف ۲.۲

کارایی کد η به صورت زیر تعریف می‌شود

$$\eta=\frac{H(U)}{L\log r}, \qquad (۱۳.۲)$$

که در آن $H(U)$ مقدار اطلاع منبع، L طول متوسط کدواژه و r حجم الفبای کد می‌باشد. ●

مثال ۵.۲

یک منبع اطلاع دارای الفبای منبع خروجی با چهار نماد u_1، u_2، u_3 و u_4 می‌باشد. الفبای کد شامل دو نماد ۰ و ۱ می‌باشد. احتمالهای نمادهای پیام همگی برابر $\frac{۱}{۴}$ هستند. فرض کنید که به صورت زیر است:

نماد	کد $(r=۲)$
u_1	۰۰
u_2	۰۱
u_3	۱۰
u_4	۱۱

برای این کد داریم: $H(U)=\log ۴=۲$ بیت، $r=۲$ و $L=۲$. از این رو کارایی کد برابر است با

$$\eta=\frac{H(U)}{L\log r}=\frac{۲}{۲\times ۱}=۱=۱۰۰\%. \qquad \blacktriangle$$

مثال ۶.۲

برای همان منبع، اکنون احتمالهای نمادهای منبع خروجی برابرند با $\frac{۱}{۲}$، $\frac{۱}{۴}$، $\frac{۱}{۸}$ و $\frac{۱}{۸}$. از طرفی برای همان کد داریم

$$H(U)=-\frac{۱}{۲}\log\frac{۱}{۲}-\frac{۱}{۴}\log\frac{۱}{۴}-۲\times\frac{۱}{۸}\log\frac{۱}{۸}=\frac{۷}{۴}\ \text{بیت} \quad ، \quad r=۲ \quad \text{و} \quad L=۲$$

در نتیجه کارایی برابر است با

$$\eta=\frac{\frac{۷}{۴}}{۲\times ۱}=\frac{۷}{۸}=۸۷/۵\%. \qquad \blacktriangle$$

مثال ۷.۲

اکنون همان منبع اطلاع کد نمایش داده شده در جدول زیر را به کار می‌برد

نماد	کد $(r=2)$
u_1	۰
u_2	۱۰
u_3	۱۱۰
u_4	۱۱۱

مجدداً داریم: $H(U)=\frac{7}{4}$ و $r=2$؛ ولی اکنون داریم

$$L=p_1 1+p_2 2+p_3 3+p_4 3=\frac{1}{2}1+\frac{1}{4}2+\frac{1}{8}3+\frac{1}{8}3=\frac{7}{4}.$$

و بنابراین کارایی برابر است با

$$\eta=\frac{\frac{7}{4}}{\frac{7}{4}\times 1}=100\%.$$

توجّه کنید که شرط کد بهینه واقعاً برقرار است: $p_i=r^{-l_i}$ برای تمام iها. ▲

۳.۲ استراتژی کدگذاری

در بخش قبل شرایط برای وجود کدهای بهینه بررسی شد تا کدهایی با کارایی بالا به دست آید. استراتژیهای گوناگونی برای یافتن کدی که به این بهینه نزدیک شود شناخته شده‌اند. در سه مثال اوّلی که از این پس ارائه خواهد شد نمادهای خروجی منبع به ترتیب احتمال نزولی قرار داده شده‌اند.

I کد فانو[1]

فانو پس از قرار دادن نمادهای منبع خروجی به ترتیب احتمال نزولی نمادها را در حدّ معقولی به r گروه با احتمالهای برابر تقسیم می‌کند. هر گروه یکی از r کدنمادها را به عنوان اوّلین نماد دریافت می‌کند. این تقسیم برای هر گروه به تعداد دفعاتی که این کار ممکن باشد تکرار می‌شود.

1. Fano

مثال ۸.۲

نماد	احتمال	$(r=۲)$ کد دودویی
u_1	۱/۴	۰۰
u_2	۱/۴	۰۱
u_3	۱/۸	۱۰۰
u_4	۱/۸	۱۰۱
u_5	۱/۱۶	۱۱۰۰
u_6	۱/۱۶	۱۱۰۱
u_7	۱/۳۲	۱۱۱۰۰
u_8	۱/۳۲	۱۱۱۰۱
u_9	۱/۳۲	۱۱۱۱۰
u_{10}	۱/۳۲	۱۱۱۱۱

▲

در مثال (۸.۲) حالتی است که $H(U) = L =$ ۲٫۸۸ و از این رو کارایی کد برابر است با $\eta =$ ۱. با وجود این، همواره ممکن نیست دقیقاً احتمالها را به گروههایی با احتمال یکسان تجزیه کرد. در این حالت، تقسیم بایستی در حدّ امکان خوب باشد. نتیجه می‌تواند به صورتهای متعدّدی ظاهر شود.

مثال ۹.۲

نماد	احتمال	$(r=۱)$ کد ۱	$(r=۲)$ کد ۲
u_1	۱/۳	۰۰	۰۰
u_2	۱/۳	۰۱	۱۰
u_3	۱/۹	۱۰	۱۱۰
u_4	۱/۹	۱۱۰	۱۱۱۰
u_5	۱/۹	۱۱۱	۱۱۱۱

▲

در مثال (۹.۲) متوسط طولهای کدواژه یکسان خواهد بود ($L =$ ۲٫۲۲). اگر چنین حالتی نباشد کد با بهترین کارایی به عنوان یک واقعیت بهتر است. کدگذاری فانو در حالتی که ۲ $< r$ نیز قابل استفاده می‌باشد.

بنابر قضیهٔ کدگذاری منبع (قضیه (۲.۲))، به طور کلّی متوسط طول کدواژه با افزایش حجم r الفبا کاهش خواهد یافت. مثال (۲. ۱۰) را ببینید.

مثال ۲. ۱۰

نماد	احتمال	کد۱ $(r=2)$	کد۲ $(r=3)$	کد۳ $(r=4)$
u_1	۰٫۳۰	۰۰	۰	۰
u_2	۰٫۲۵	۰۱	۱۰	۱
u_3	۰٫۱۲	۱۰۰	۱۱	۲۰
u_4	۰٫۱۰	۱۰۱	۲۰	۲۱
u_5	۰٫۱۰	۱۱۰	۲۱	۳۰
u_6	۰٫۰۵	۱۱۱۰	۲۲۰	۳۱
u_7	۰٫۰۴	۱۱۱۱۰	۲۲۱	۳۲
u_8	۰٫۰۴	۱۱۱۱۱	۲۲۲	۳۳
		$L_1 = 2{,}66$	$L_2 = 1{,}83$	$L_3 = 1{,}45$
		$\eta_1 = 0{,}99$	$\eta_2 = 0{,}91$	$\eta_3 = 0{,}91$

$H(U) = 2{,}64$ نماد / بیت ▲

II کد شانون

شانون یک سری از احتمالهای تجمعی $p_k = \sum_{i=1}^{k-1} p(u_i)$ برای $k = 1,2,\ldots,n$ را حساب کرد. اینها را بعداً (برای کد دودویی) به صورت دودویی می‌نویسیم. تعداد نمادها در هر کدواژه از نابرابری زیر به دست می‌آید

$$\log\frac{1}{p_k} \le l_k < \log\frac{1}{p_k} + 1.$$

در مثال زیر u_6 نماد منبع خروجی با ۱۱۰۱ نمایش داده شده است، چون $p_6 = \frac{13}{16}$ که می‌توان آن را به صورت $1\times 2^{-1} + 1\times 2^{-2} + 0\times 2^{-3} + 1\times 2^{-4}$ نوشت. تعداد نمادهای کد باید حداقل برابر $\log(\frac{1}{p_k}) = \log 16 = 4$ باشد، به طوری که برای این نماد نیازی نیست هیچ صفری به کد اضافه کرد.

مثال ۱۱.۲

نماد	احتمال	P_i	طول l_i	کد $(r=۲)$
$u_۱$	۱/۴	$p_۱ = ۰$	$l_۱ = ۲$	$= ۰۰$
$u_۲$	۱/۴	$p_۲ = ۱/۴$	$l_۲ = ۲$	$= ۰۱$
$u_۳$	۱/۸	$p_۳ = ۱/۲$	$l_۳ = ۳$	$= ۱۰۰$
$u_۴$	۱/۸	$p_۴ = ۵/۸$	$l_۴ = ۳$	$= ۱۰۱$
$u_۵$	۱/۱۶	$p_۵ = ۳/۴$	$l_۵ = ۴$	$= ۱۱۰۰$
$u_۶$	۱/۱۶	$p_۶ = ۱۳/۱۶$	$l_۶ = ۴$	$= ۱۱۰۱$
$u_۷$	۱/۳۲	$p_۷ = ۷/۸$	$l_۷ = ۵$	$= ۱۱۱۰۰$
$u_۸$	۱/۳۲	$p_۸ = ۲۹/۳۲$	$l_۸ = ۵$	$= ۱۱۱۰۱$
$u_۹$	۱/۳۲	$p_۹ = ۱۵/۱۶$	$l_۹ = ۵$	$= ۱۱۱۱۰$
$u_{۱۰}$	۱/۳۲	$p_{۱۰} = ۳۱/۳۲$	$l_{۱۰} = ۵$	$= ۱۱۱۱۱$

▲

از مثال (۱۱.۲) ممکن است نتیجه‌گیری شود که در این حالت روش شانون به همان کدی منجر می‌شود که از روش فانو به دست آمد (با مثال (۸.۲) مقایسه کنید)؛ در مثال زیر نشان داده شده است که همواره این چنین نیست.

مثال ۱۲.۲

نماد	احتمال	P_i	l_i	کد شانون $(r=۲)$	کد فانو $(r=۲)$
$u_۱$	۰٫۴	۰	۲	۰۰	۰
$u_۲$	۰٫۳	۰٫۴	۲	۰۱	۱۰
$u_۳$	۰٫۲	۰٫۷	۳	۱۰۱	۱۱۰
$u_۴$	۰٫۱	۰٫۹	۴	۱۱۱۰	۱۱۱

▲

نتیجهٔ خاصّ کدگذاری فانو در مثال (۱۲.۲) را ***کد کاما***[1] نیز می‌نامند، زیرا عدد دودویی ۰ پایان یک کدواژه را نشان می‌دهد و علاوه بر آن هیچ کدواژه‌ای طولی بیشتر از ۳ ندارد.

1. Comma Cod

III کد هافمن[1]

در کد هافمن در حالت دودویی دو نماد منبع خروجی با کمترین احتمال با یکدیگر پیوند زده می‌شوند و یک الفبای پیام جدید با یک نماد کمتر نتیجه می‌شود. در الفبای جدید پس از آن که دو نماد دوباره با هم پیوند زده شدند الفبای جدیدی ارائه می‌شود. این کار تا زمانی که الفبای پیام درست به دو نماد منجر شود ادامه داده می‌شود. به این دو نماد نمادهای ۰ و ۱ از کد دودویی نسبت داده می‌شود. با حرکت به عقب در هر محلّی که دو نماد با هم پیوند زده شده ۰ یا ۱ به کدواژه اضافه می‌شود.

مثال ۱۳.۲

نماد	احتمال					کد $(r=2)$
u_1	۰٫۴	۰٫۴	۰٫۴	۰٫۴	۰٫۶ (۰)	۱
u_2	۰٫۳	۰٫۳	۰٫۳	۰٫۳ (۰)	۰٫۴ (۱)	۰۰
u_3	۰٫۱	۰٫۱	۰٫۲ (۰)	۰٫۳ (۱)		۰۱۱
u_4	۰٫۱	۰٫۱ (۰)	۰٫۱ (۱)			۰۱۰۰
u_5	۰٫۰۶ (۰)	۰٫۱ (۱)				۰۱۰۱۰
u_6	۰٫۰۴ (۱)					۰۱۰۱۱

▲

مثال داده شده نتیجه‌ای خواهد داد که با روش فانو قابل مقایسه است؛ روش شانون کد با کارایی کمتری را می‌دهد. به طور کلّی، روش هافمن منجر به کد کاراتری می‌شود.

اگر تعداد کدنمادها r باشد، در این صورت برای یک کد هافمن بهینه باید $r+k(r-1)$ نماد منبع خروجی (k عدد صحیح) وجود داشته باشد. اگر کدی نمادهای منبع خروجی کمتری داشته باشد در این صورت باید نمادهایی اضافه نموده و احتمال آنها را برابر صفر قرار دهیم. مثال زیر را ملاحظه نمایید.

مثال ۱۴.۲

برای متوسط طول کدواژه در حالت زیر داریم: $L=1{,}67$. اگر نماد u_7 اضافه نشود و کد هافمن به کار برده شود در این صورت نتیجه خواهد شد $L=2$.

1. Huffman

نماد	احتمال			کد $(r=3)$
u_1	۱/۳	۱/۳	۴/۹ (۰)	۱
u_2	۱/۶	۲/۹	۱/۳ (۱)	۰۰
u_3	۱/۶	۱/۶ (۰)	۲/۹ (۲)	۰۱
u_4	۱/۹	۱/۶ (۱)		۰۲
u_5	۱/۹ (۰)	۱/۹ (۲)		۲۰
u_6	۱/۹ (۱)			۲۱
u_7	۰ (۲)			۲۲

▲

IV کد ژیلبرت[1]-مور[2] (کد الفبایی)

یک روش کاملاً متفاوت توسط کد ژیلبرت-مور داده می‌شود. با این روش نمادهای منبع خروجی را می‌توان به هر ترتیب دلخواهی قرار داد (مثلاً، الفبایی). اگر طول کدواژه u_i با l_i داده شده باشد، در این صورت این طول توسط نابرابری زیر تعیین می‌شود

$$2^{1-l_i} \le p(u_i) < 2^{2-l_i}\ ,\quad i=1,2,\ldots,n. \tag{۱۵.۲}$$

سپس سری نا-کاهشی $(\alpha_1,\alpha_2,\ldots)$ به صورت زیر تعیین می‌شود

$$\left.\begin{aligned}
\alpha_1 &= \frac{1}{2}p(u_1)\\
\alpha_2 &= p(u_1)+\frac{1}{2}p(u_2)\\
&\vdots\\
\alpha_i &= p(u_1)+p(u_2)+\cdots+p(u_{i-1})+\frac{1}{2}p(u_i)
\end{aligned}\right\} \tag{۱۶.۲}$$

در این جا داریم: $0 \le \alpha_1 \le \alpha_2 \cdots \le 1$. کد برای پیام u_i با نمایش عدد α_i به صورت سری دودویی با طول l_i به دست می‌آید.

مثال ۱۵.۲

سه حرف اوّل الفبا a، b و c با احتمالهای معلوم رخ‌داد برای زبان انگلیسی را (شکل (۲.۲) را ببینید) در نظر بگیرید؛ داریم

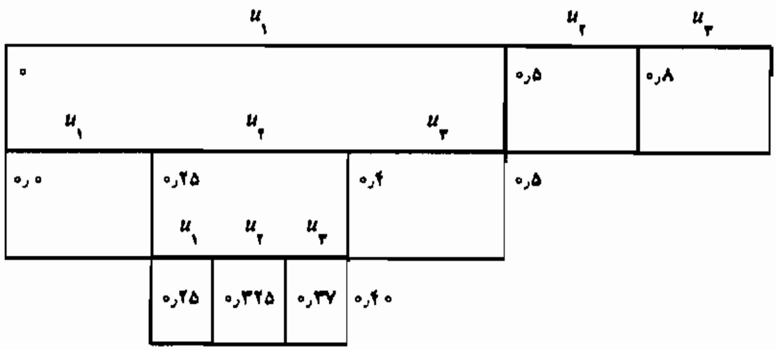
1. Gilbert 2. Moore

نماد	احتمال	l_i	α_i	کد ($r=۲$)
$u_۱$	۰٫۰۶۴	۵	۰٫۰۳۲	۰۰۰۰۱
$u_۲$	۰٫۰۱۳	۸	۰٫۰۷۱	۰۰۰۱۰۰۱۰
$u_۳$	۰٫۰۲۲	۷	۰٫۰۸۸	۰۰۰۱۰۱۱۱

▲

۷ کد حسابی

با این کد احتمالهای $p(u_۱)$، $p(u_۲)$، ...، $p(u_n)$ نمادهای منبع خروجی به عنوان زیرفاصله‌های فاصلهٔ واحد [۰,۱] نمایش داده شده‌اند (مجموع احتمالها برابر یک است). ماهیت این روش را برای حالتی که $n=۳$، با $p(u_۱)=۰٫۵$، $p(u_۲)=۰٫۳$ و $p(u_۳)=۰٫۲$ بررسی خواهیم کرد. احتمالهای تجمعی به ترتیب برابرند با $P_۱=۰$، $P_۲=۰٫۵$ و $P_۳=۰٫۸$. این مطلب در شکل (۳.۲) نشان داده شده است. هر نقطه احتمال تجمعی P_i نمادهای $u_۱$، $u_۲$، ...، $u_{i-۱}$ را نشان می‌دهد.



شکل ۳.۲- کدگذاری دنبالهٔ $u_۱$، $u_۲$، ... به کمک یک کد حسابی

اگر نماد $u_۱$ منبع خروجی رخ داده باشد (با احتمال ۰٫۵) این با فاصلهٔ [۰٫۰,۰٫۵] متناظر است. بعداً فرض می کنیم که نماد دومی توسط منبع تولید شده است. اکنون فاصلهٔ جاری یعنی [۰٫۰,۰٫۵] مجدداً برطبق توزیع احتمال تجمعی منبع یعنی ۰٫۰، ۰٫۵، ۰٫۸ و ۱٫۰ به زیرفاصله‌هایی تقسیم می‌شود. اکنون سه زیرفاصلهٔ [۰٫۰,۰٫۲۵]، [۰٫۲۵,۰٫۴۰] و [۰٫۴۰,۰٫۵۰] ملاحظه می‌شود. حال اگر نماد $u_۲$ تولید شده باشد در این صورت بعد از دو نماد منبع در فاصلهٔ [۰٫۲۵,۰٫۴۰] هستیم. با تکرار این فرایند، نتیجه‌ای ارائه می‌دهیم که یک دنباله از نمادهای منبع توسط یک زیرفاصله [۰,۱] نمایش داده می‌شود که به طور یکتا با آن جفت شده است. اکنون نمایش دودویی چپ‌ترین نقطهٔ فاصله را به عنوان کدواژه متعلّق به دنبالهٔ نمادهای

منبع u_1، u_2، ... در نظر می‌گیریم و الی‌آخر. پهنای فاصله را متناظر با احتمال رخ‌داد دنبالهٔ نظیر نمادهای منبع در نظر می‌گیریم.

اکنون می‌توان کد را به صورت فرایند بازگشتی که به صورت زیر عمل می‌کند در نظر گرفت. برای هر مرحله، یعنی وقتی مجدداً یک نماد جدید منبع معرّفی می‌شود، چپ‌ترین نقطهٔ C از فاصلهٔ جاری و پهنای جاری A از این فاصله را در نظر می‌گیریم. چپ‌ترین نقطهٔ جدید برابر با چپ‌ترین نقطهٔ قدیم به اضافهٔ قسمتی از پهنای فاصلهٔ جاری است، برطبق رابطهٔ زیر

$$C_{\text{جدید}} = C_{\text{قدیم}} + A_{\text{قدیم}}\ P_i \qquad (۱۷.۲)$$

که در آن P_i احتمال تجمعی برای نماد u_i می‌باشد. پهنای فاصلهٔ جدید با ضرب پهنای قدیم در احتمال p_i به دست می‌آید؛ از این رو داریم:

$$A_{\text{جدید}} = A_{\text{قدیم}}\ P_i \qquad (۱۸.۲)$$

در مثال شکل (۳.۲) در شروع داریم (مرحله ۰ام):

$$C_{\text{شروع}} = ۰٫۰$$

$$A_{\text{شروع}} = ۱٫۰$$

پس از u_1 نتایج زیر را برای C و A داریم:

$$C_{\text{جدید}} = ۰٫۰ + ۱٫۰ \times ۰٫۰ = ۰٫۰$$

$$A_{\text{جدید}} = ۱٫۰ \times ۰٫۵ = ۰٫۵$$

پس از نماد دوم، u_2 در این مثال، داریم:

$$C_{\text{جدید}} = ۰٫۰ + ۰٫۵ \times ۰٫۵ = ۰٫۲۵$$

$$A_{\text{جدید}} = ۰٫۵ \times ۰٫۳ = ۰٫۱۵$$

اگر پیوسته یک گروه l نمادی را برطبق این مسیر بازگشتی کدگذاری کنیم یک نقطهٔ چپ C و پهنای A از فاصلهٔ حاصل به دست می‌آوریم. برای این که کدواژهٔ نهایی را بیابیم نقطهٔ چپ C را انتخاب می‌کنیم و آن را به شکل دودویی با تعدادی بیت می‌نویسیم به قسمی که بتوان آن را از نقاط چپ فاصله‌های دیگر تشخیص داد. از این رو در مثال جفت نماد منبع u_1u_2 به عنوان نمایش دودویی عدد اعشاری ۰٫۲۵ و به صورت دودویی ۰٫۰۱ کدگذاری شده است. در حقیقت، کدگشایی فرایند معکوس را با تعیین کردن مرحله به

مرحله از طریق نمادی که فاصلهٔ جاری را پدید آورده دنبال می‌کنیم، و فاصلـهٔ قبلـی را از روی این فاصله تعیین می‌کنیم.

VI کدگذاری براساس توسعهٔ الفبا

گرچه در این فصل منبع بی‌حافظه در نظر گرفته شده بنابراین نمادهای متوالـی منبـع مستقلّند، با این وجود کدگذاری ترکیبی از نمادها می‌تواند به کـد کـارایی منجـر شـود. بنابراین l نماد منبع را به یک پیام دسته‌بندی می‌کنیـم و احتمـال ایـن پیامهـا را محاسـبه می‌کنیم و سپس یک استراتژی کدگذاری نظیر هافمن را برای تعیین کد احتمالی به کـار می‌بریم. این روش را که به عنوان توسعهٔ الفبا شناخته شده است با کمک یک مثال بررسی خواهیم کرد.

مثال ۱۶.۲

دو نماد u_1 و u_2 منبع خروجی دارای احتمالهای $\frac{۳}{۴}$ و $\frac{۱}{۴}$ می‌باشند. کدگذاری بر مبنـای روش کدگذاری فانو انجام شده است.

نماد	احتمال	کد $(r=2)$
u_1	۳/۴	۰
u_2	۱/۴	۱

با $\eta=$ ۰٫۸۱۱ و $r=$ ۲ و $L=$ ۱ و بیت $H(U)=$ ۰٫۸۱۱.

سپس دو نماد را با هم اختیار می‌کنیم و از این رو پیام جدیـد v_1، ...، v_4 بـه دسـت می‌آوریم.

اکنون داریم

$$\eta=0.961 ،\ r=2 ،\ L=27.16 ،\ H(V)=1.622$$

پیام	احتمال	کد $(r=2)$
$v_1 = u_1 u_1$	$p(v_1)=p(u_1,u_1)=$ ۹/۱۶	۰
$v_2 = u_1 u_2$	$p(v_2)=p(u_1,u_2)=$ ۳/۱۶	۱۰
$v_3 = u_2 u_1$	$p(v_3)=p(u_2,u_1)=$ ۳/۱۶	۱۱۰
$v_4 = u_2 u_2$	$p(v_4)=p(u_2,u_2)=$ ۱/۱۶	۱۱۱

از این رو با اختیار دو نماد منبع با همدیگر کارایی افزایش یافته است. ▲

با کدگذاری l نماد منبع با همدیگر منبع جدیدی با یک الفبای توسعه یافته یعنی با n^l پیام (به جای منبع اصلی با n نماد) نتیجه می‌شود، که در آن احتمال یک پیام با حاصل‌ضرب احتمالها نمادی که پیام را می‌سازند برابر است.

۴.۲ محتملترین پیامها

در بخش (۱.۲) مجموعهٔ $V = \{v_1, v_2, \ldots, v_j, \ldots, v_{n^l}\}$ را با پیامهای v شامل l نماد منبع از یک منبع الفبا با اندازهٔ n معرّفی کردیم. با افزایش l به نظر می‌رسد که برخی از پیامها دارای احتمال رخ‌داد قابل صرف‌نظر کردن باشند در حالی که بقیه تقریباً احتمال برابر دارند. در این صورت می‌توان از تعداد محتملترین پیامها سخن گفت که البته کمتر از تعداد پیامهای ممکن می‌باشد.

فرض کنید ℓ_i $(\ell_i > 0)$ تعداد دفعاتی باشد که u_i نماد منبع در پیام v_j ظاهر می‌شود. در این صورت در حالت منبع بی‌حافظه احتمال یک پیام دلخواه v با رابطهٔ زیر داده می‌شود

$$p(v) = \prod_{i=1}^{k} p(u_i)^{\ell_i}, \tag{۱۹.۲}$$

که در آن k تعداد نمادهای متفاوت در پیام v می‌باشد و داریم

$$\ell = \sum_{i=1}^{k} \ell_i. \tag{۲۰.۲}$$

و از این رو

$$\log p(v) = \log\{\prod_{i=1}^{k} p(u_i)^{\ell_i}\}. \tag{۲۱.۲}$$

اگر بنا به قانون اعداد بزرگ $\ell_i / \ell \rightarrow p(u_i)$، یعنی $\ell_i \approx \ell p(u_i)$، داریم

$$\begin{aligned} \log p(v) &\approx \log\{\prod_{i=1}^{k} p(u_i)^{\ell p(u_i)}\} \\ &= \ell \sum_{i=1}^{k} p(u_i) \log p(u_i) \\ &= -\ell H(U). \end{aligned} \tag{۲۲.۲}$$

از این رو

$$\frac{1}{\ell}\log p(v) \approx -H(U). \qquad (۲۳.۲)$$

یعنی $H(U)$ تقریباً برابر است با لگاریتم وارون احتمال یک دنباله معمولاً طولانی بخش بر تعداد نمادها در دنباله. این موضوع برای هر منبعی درست است. به طور دقیق‌تر قضیهٔ زیر را داریم.

قضیهٔ ۳.۲ (قضیهٔ شانون- مک‌میلان)

یک منبع بی‌حافظهٔ گسسته با الفبای U و اطلاع $H(U)$ داده شده است. در این صورت با هر $\varepsilon > 0$ و $\delta > 0$ داده شده می‌توان ℓ_0 را به قسمی یافت که واژه‌های منبع به طول $\ell \geq \ell_0$ در دو دسته قرار گیرند.

(الف) مجموعهٔ S' که احتمالش کمتر از ε است.

(ب) مجموعهٔ باقی‌ماندهٔ S، تمام اعضایی که دارای احتمالهایی هستند که در نابرابری زیر صدق می‌کنند

$$\left|\frac{-\log p(v)}{\ell} - H(U)\right| \leq \delta. \qquad (۲۴.۲)$$

برهان

مجموعهٔ S را می‌توان به صورت زیر تعریف کرد

$$S = \{v \,|\, |-\log p(v) - \ell H(U)| < \ell\delta\}. \qquad (۲۵.۲)$$

نابرابری چبیشف بیان می‌کند که برای هر متغیّر تصادفی x با میانگین μ و واریانس σ^2

$$p\{|x-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}. \qquad (۲۶.۲)$$

با به کارگیری این نابرابری دربارهٔ مجموعهٔ S به دست می‌آوریم

$$p\{|-\log p(v) - \ell H(U)| \geq \ell\delta\} \leq \frac{\mathrm{var}[-\log p(v)]}{\ell^2\delta^2}$$

$$\leq \frac{\ell\delta^2}{\ell^2\delta^2} = \frac{\delta^2}{\ell\delta^2}, \qquad (۲۷.۲)$$

که در آن

$$\sigma^{2} = \sum_{p(u_i)} p(u_i)\big(\log p(u_i)\big)^{2} - \left(\sum_{p(u_i)} p(u_i)\log p(u_i)\right)^{2}$$

ثابتی مستقل از ℓ است.

در نتیجه برای ℓ به قدر کافی بزرگ $p(S')$ احتمال این‌که کدواژه‌ای که رخ می‌دهد متعلّق به مجموعهٔ S' باشد از ε کمتر است . ■

واضح است، $p(S)=\sum_{v\in S} p(v)$ احتمال این‌که کدواژه‌ای که رخ می‌دهد متعلّق به مجموعهٔ S باشد در نابرابری زیر صدق می‌کند

$$1-\varepsilon < p(S) \le 1. \qquad (۲۸.۲)$$

برای ε کوچک S شامل تمام کدواژه‌هایی است که با احتمال بالایی رخ می‌دهند. به این دلیل S را مجموعهٔ ***محتملترین واژه‌های منبع*** یا ***محتملترین پیامها*** نیز می‌نامند. قضیهٔ زیر کرانهایی برای تعداد اعضای S ارائه می‌کند که با $M=|S|$ نشان داده می‌شود.

قضیهٔ ۴.۲

برای M تعداد کدواژه‌های منبع در S مجموعهٔ محتملترین واژه‌های منبع، داریم

$$(1-\varepsilon)2^{\ell\{H(U)-\delta\}} \le M \le 2^{\ell\{H(U)+\delta\}}. \qquad (۲۹.۲)$$

برهان

از

$$\left|\frac{-\log p(v)}{n} - H(U)\right| < \delta$$

نتیجه می‌شود که

$$2^{-\ell\{H(U)+\delta\}} \le p(v) \le 2^{-\ell\{H(U)-\delta\}} \qquad (۳۰.۲)$$

و از این رو داریم

$$\sum_{v\in S} 2^{-\ell\{H(U)+\delta\}} < p(S) \le \sum_{v\in S} 2^{-\ell\{H(U)-\delta\}}, \qquad (۳۱.۲)$$

یا

$$M2^{-\ell\{H(U)+\delta\}} < p(S) \le M2^{-\ell\{H(U)-\delta\}}. \qquad (۳۲.۲)$$

همچنین (معادلهٔ (۲۸.۲) را ببینید)

$$1-\varepsilon \le p(S) \le 1.$$

بنابراین

$$1-\varepsilon \le M 2^{-\ell\{H(U)-\delta\}} \tag{۳۳.۲}$$

و

$$M 2^{-\ell\{H(U)+\delta\}} \le 1 \tag{۳۴.۲}$$

که قضیه را ثابت می‌کند. ■

برای ε و δ کوچک نتیجه می‌شود که

$$M \approx 2^{-\ell H(u)}. \tag{۳۵.۲}$$

چون مقدار اطلاع منبع حداکثر برابر با $\log n$ است، یعنی اگر همهٔ واژه‌های منبع احتمال یکسان داشته باشند، نتیجه می‌شود که

$$M_{\max} = 2^{\ell \log n} = n^{\ell}, \tag{۳۶.۲}$$

که دقیقاً برابر تعداد پیامهای ممکن است. بنابراین برای منبعی کـه در آن $H(U) < \log n$، تعداد محتملترین پیامها با طول ℓ کوچکتر از تعداد پیامهای ممکن خواهد بود. همچنین می‌توان گفت که اگر طول ℓ بزرگ باشد بخشی از پیامهای ممکن دارای احتمال رخ‌داد کوچک قابل صرف‌نظر کردن می‌باشند.

همان طور که در معادلهٔ (۳۵.۲) بیان شد یک رابطهٔ نمایی بین تعداد محتملترین پیامهای یک منبع اطلاع گسسته M و حجم اطلاع منبع وجود دارد. برای منبعی که ماکسیمم مقدار اطلاع تولید می‌کند تعداد پیامهای محتمل با تعداد پیامهای ممکن برابر است. بنابراین می‌توان مقدار اطلاع را به عنوان اندازهٔ تعداد پیامهایی که یک منبع دقیقاً می‌تواند تولید کند (یعنی، با احتمالی معلوم) در نظر گرفت.

به کمک رهیافتی که در این جا معرّفی شد ممکن است مطالبی را که در بخشهای (۲.۲) و (۳.۲) بررسی شد بیشتر تشریح نمود. این کار را در این جا برای روشن شدن این که چگونه مفهوم «تعداد محتملترین پیامها» ممکن است به کار برده شود، انجام خواهیم داد. فرض می‌کنیم که پیامهای منبع دارای طول ℓ هستند و کدواژه‌ای به طول L به هر پیام منبع نسبت می‌دهیم. نمادهای منبع از یک الفبای $U = \{u_1, u_2, \ldots, u_n\}$ و کدواژه‌ها از الفبای $S = \{s_1, s_2, \ldots, s_r\}$ انتخاب شده‌اند.

در این صورت تعداد پیامهای ممکن منبع برابر n^{ℓ} است و تعداد کدواژه‌های ممکن

موجود برابر r^L است. بنابراین ممکن است هر پیام را به یک کدواژه‌ای کدگذاری کرد اگر

$$r^L \geq n^\ell$$

یا

$$\frac{L}{\ell} \geq \frac{\log n}{\log r} = \log_r n,$$

چون در این صورت یک کدواژه برای هر پیام منبع وجود دارد. با وجود این، نتیجه می‌شود که تنها باید M تعداد محتملترین پیامها را مورد توجّه قرار داد. این امر به قضیهٔ کدگذاری منبع زیر منجر می‌شود.

قضیهٔ ۵.۲ (قضیهٔ اوّل کدگذاری شانون)

منبع اطلاع گسسته بی‌حافظه‌ای با مقدار اطلاع $H(U)$ که از آن پیامهایی با طول ℓ با کدواژه‌هایی به طول L از یک الفبای کد با حجم r کدگذاری شده‌اند مفروض است. اگر P_e احتمال آن باشد که پیامی رخ می‌دهد که برای آن هیچ کدواژه‌ای وجود ندارد، در این صورت می‌توان P_e را به دلخواه کوچک کرد. مثلاً $P_e < \varepsilon$، تا زمانی که طول L در

$$L \log r \geq \ell\, H(U)$$

صدق کند و ℓ به قدر کافی بزرگ باشد.

برهان

از قضیهٔ قبل می‌دانیم که تعداد واژه‌های منبع در S از $۲^{\ell\{H(U)+\delta\}}$ کمتر است. با فرض قضیه داریم

$$L \log r \geq \ell\, H(U). \tag{۳۷.۲}$$

بنابراین δای می‌توان انتخاب کرد به قسمی که

$$L \log r \geq \ell\{H(U)+\delta\}, \tag{۳۸.۲}$$

و

$$r^L \geq ۲^{\ell\{H(U)+\delta\}}. \tag{۳۹.۲}$$

در نتیجه تعداد کدواژه‌ها بلوکی به طول L ($= r^L$) از تعداد واژه‌های منبع در S بزرگتر است. مجموعه‌ای که برای آن کدواژه‌ای وجود ندارد در S' قرار دارد و برای ℓ به قدر کافی بزرگ داریم، $P_e < \varepsilon$ همان طور که باید ثابت می‌شد. ■

این بدین معنی است که وقتی یک واژهٔ منبع با طول زیاد انتخاب شده باشد می‌توان بی-خطا کدگذاری کرد، حتّی اگر تعداد کدواژه‌های ممکن از تعداد پیامهای ممکن کمتر باشد. در واقع، به این دلیل است که پیوند نمادهای منبع همان طور که در بخش (۳.۲) تذکّر داده شد می‌تواند به کد کاراتری منجر گردد.

۵.۲ تمرینها

در تمرینهای ذیل فرض شده است که نمادهای متوالی به طور آماری مستقلّند.

۱.۲ یک منبع اطلاع نمادهایی که متعلّق به الفبای $U = \{u_1, u_2, u_3\}$ است تولید می‌کند. احتمالهای این نمادها به ترتیب برابر است با ۷ر۰، ۲ر۰ و ۱ر۰.

(الف) مقدار اطلاع هر نماد را حساب کنید.

(ب) احتمالهای تمام پیامهای ممکن شامل دو نماد را حساب کنید.

(پ) مقدار اطلاع را برای هر دو بند (الف) و (ب) برای هر پیام دونمادی حساب کنید.

(ت) حشو منبع اطلاع را حساب کنید.

(ث) مقدار اطلاع بر ثانیه را حساب کنید، اگر معلوم باشد که مدّت u_1 برابر ۰۰۱ر۰ ثانیه، مدّت u_2 برابر ۰۰۲ر۰ ثانیه، و مدّت u_3 برابر ۰۰۳ر۰ ثانیه است.

۲.۲ یک منبع اطلاع هشت نماد مختلف (u_1 تا u_8) را به ترتیب با احتمالهای $\frac{۱}{۲}$، $\frac{۱}{۴}$، $\frac{۱}{۸}$، $\frac{۱}{۱۶}$، $\frac{۱}{۳۲}$، $\frac{۱}{۶۴}$، $\frac{۱}{۱۲۸}$ و $\frac{۱}{۱۲۸}$ تولید می‌کند. این نمادها به ترتیب به صورت ۰۰۰، ۰۰۱، ۰۱۰، ۰۱۱، ۱۰۰، ۱۰۱، ۱۱۰ و ۱۱۱ کدگذاری شده‌اند.

(الف) مقدار اطلاع برای هر نماد چه‌قدر است؟

(ب) احتمالهای رخداد برای یک ۰ و یک ۱ چه‌قدر است؟

(پ) کارایی این کد چیست؟

(ت) با کمک روش فانو یا شانون یک کد کارا بدهید.

(ث) کارایی کدی که به این طریق به دست آمده چه‌قدر است؟

۳.۲ شش نماد پیام ($u_1,\ldots,u_6$) که با احتمالهای $\frac{۳}{۸}$، $\frac{۱}{۶}$، $\frac{۱}{۸}$، $\frac{۱}{۸}$، $\frac{۱}{۸}$ و $\frac{۱}{۱۲}$ رخ می‌دهند با استفاده از کد سه تایی (نمادهای کد ۰، ۱ و ۲) باید کدگذاری شوند.

(الف) کد خواسته شده را با بهره‌وری از روش فانو تعیین کنید.

(ب) کارایی کد به دست آمده را تعیین کنید.

۴.۲ هفت نماد پیام $(u_1,...,u_7)$ که با احتمالهای $\frac{1}{4}$، $\frac{1}{5}$، $\frac{1}{6}$، $\frac{1}{6}$، $\frac{1}{12}$، $\frac{1}{12}$ و $\frac{1}{20}$ رخ می‌دهند باید با چهار کدنماد a، b، c و d کدگذاری شوند.

(الف) کد مورد نیاز را با بهره‌وری از روش فانو به دست آورید.
(ب) کارایی کد به دست آمده را تعیین کنید.

۵.۲ یک منبع اطلاع دارای الفبایی شامل پنج نماد $(u_1,...,u_5)$ است. احتمال رخ‌داد این نمادها به ترتیب عبارتند از $\frac{1}{2}$، $\frac{1}{4}$، $\frac{1}{8}$، $\frac{1}{16}$ و $\frac{1}{16}$.

(الف) کد مناسبی برای این نمادهای پیام با سه کدنماد a، b و c تعیین کنید.
(ب) کارایی کد به دست آمده را تعیین کنید.

۶.۲ یک منبع اطلاع دارای الفبایی شامل سه نماد (a,b,c) است. احتمال رخ‌داد این نمادها به ترتیب عبارتند از $\frac{1}{2}$، $\frac{1}{3}$ و $\frac{1}{6}$.

(الف) کد مناسب دودویی برای این نمادهای پیام بیابید.
(ب) کارایی کد حاصل را تعیین کنید.
(پ) اصلاح کارایی با انتخاب پیوستهٔ دو نماد پیام با هم مورد بررسی است. یک کد مناسب دودویی برای این کار بیابید.
(ت) کارایی کد به دست آمده در (پ) را تعیین کنید.

۷.۲ یک منبع اطلاع دارای الفبای منبع از ۹ نماد مختلف u_1 تا u_9 با احتمالهای به ترتیب $\frac{1}{4}$، $\frac{1}{4}$، $\frac{1}{8}$، $\frac{1}{8}$، $\frac{1}{16}$، $\frac{1}{16}$، $\frac{1}{16}$، $\frac{1}{32}$ و $\frac{1}{32}$ می‌باشد. این منبع اطلاع به یک کانال ارتباطی متصل شده است که از سه نماد a، b و c استفاده می‌کند.

(الف) کد و کارایی کد را بر مبنای روشهای فانو و هافمن بیابید.
(ب) به همین سؤال درحالتی که نماد c هرگز نمی‌تواند با c دیگری دنبال شود پاسخ دهید. توجّه کنید کد بایستی قابل کدگشایی فوری باشد.

۸.۲ در یک محلّ اخذ رأی دو سؤال مستقل پرسیده می‌شود که می‌توان هر یک را با بلی، خیر و هیچ عقیده‌ای پاسخ داد. برای یک مرکز قبلی که همین سؤالات پرسیده شده‌اند، نتایج به صورت زیر بوده است.

	بلی	خیر	هیچ عقیده
سؤال ۱	۵۰٪	۴۰٪	۱۰٪
سؤال ۲	۶۰٪	۲۰٪	۲۰٪

براساس این داده‌ها می‌خواهیم برای ترکیب دو پاسخ یک کد بسازیم و بین کد دودویی و سه‌سه‌ای مقایسه‌ای انجام دهیم.

(الف) یک کد دودویی مناسب برطبق روش فانو تعیین کنید.
(ب) یک کد سه‌سه‌ای مناسب برطبق روش فانو بیابید.
(پ) تمایل خود را بر مبنای کارایی هر دو کد بیان کنید.

۹.۲ یک کد برطبق روش ژیلبرت و مور برای u_i نماد پیام مشخّص شده به صورت زیر ارائه نمایید.

u_i	$u_۱$	$u_۲$	$u_۳$	$u_۴$
$p(u_i)$	۰٫۱	۰٫۳	۰٫۲	۰٫۴

این روش را بر مبنای کارایی با یک کدگذاری به روش شانون و یک کدگذاری به روش هافمن مقایسه کنید.

۱۰.۲ یک آزمایش دارای برآمدهای ممکن ۱، ۲، ۳، ۴، ۵، ۶ و ۷ و با احتمالهای

$$p(۱) = p(۲) = \frac{۱}{۳} \ ، P(۳) = p(۴) = \frac{۱}{۹} \ ، p(۵) = p(۶) = p(۷) = \frac{۱}{۲۷}$$

می‌باشد. می‌خواهیم برآمدهای آزمایش را با یک کانال دودویی یا سه‌سه‌ای ارسال کنیم؛ هر دو کانال بدون نوفه‌اند. هزینهٔ کانال دودویی ۱٫۸۰ پوند بر کدنماد و برای کانال سه‌سه‌ای ۲٫۷۰ پوند بر کدنماد می‌باشد.

(الف) یک کد برای کانال دودویی برطبق روش هافمن بیابید و کارایی آن را تعیین کنید.
(ب) یک کد برای کانال سه‌سه‌ای برطبق روش فانو بیابید و کارایی آن را تعیین کنید.
(پ) کدام کانال را (با کد براساس بند (الف) و (ب)) در صورتی که بخواهیم متوسط هزینه‌ها را می‌نیمم کنیم ترجیح می‌دهید؟ در این صورت مقدار این هزینه چه قدر

است؟

۱۱.۲ یک منبع اطلاع، صفر و یک را با احتمالهای $p(0) = 0{,}8$ و $p(1) = 0{,}2$ تولید می‌کنـد. دنبالهٔ نمادها با کدنمادهای ۰، ۱، ۲، ۳ و ۴ برطبق جدول زیر کدگذاری شده‌اند.

کد	پیام
۰	۱
۱	۰۱
۲	۰۰۱
۳	۰۰۰۱
۴	۰۰۰۰

(الف) آیا این کد به طور یکتا قابل کدگشایی و قابل کدگشایی فوری است؟
(ب) متوسط مقدار اطلاع برای هر کدنماد را تعیین کنید.
(پ) کارایی این کد چه قدر است؟

۱۲.۲ یک منبع، پیامها را از الفبایی که شامل هشت نماد u_1 تـا u_8 اسـت تولیـد می‌کنـد. احتمالهای رخداد عبارتند از ۳۲ر۰، ۲۴ر۰، ۲۰ر۰، ۰۹ر۰، ۰۵ر۰، ۰۴ر۰، ۰۴ر۰ و ۰۲ر۰.

(الف) یک کد مناسب دودویی با کمک روش شانون بیابید.
(ب) یک کد مناسب سه مقداری با کمک روش هافمن بیابید.
(پ) یک کد مناسب چهار مقداری با کمک روش فانو بیابید.
(ت) کارایی این کدها را مقایسه کنید.
(ث) مقدار اطلاع برای هر نماد را برای کد دودویی (با گرد کردن احتمالها تا یک رقم اعشار) بیابید.

۶.۲ جوابها

۱.۲ (الف) چون نمادهای متوالی به طور آماری مستقلّند، برای مقدار اطلاع بـرای هـر نماد به دست می‌آوریم

$$H(U) = -\sum_{i=۱}^{۳} p(u_i)\log p(u_i) = -\frac{۷}{۱۰}\log\frac{۷}{۱۰} - \frac{۲}{۱۰}\log\frac{۲}{۱۰} - \frac{۱}{۱۰}\log\frac{۱}{۱۰}$$

$$= -\frac{۷}{۱۰}\log ۷ - \frac{۲}{۱۰}\log ۱۰ = ۱{,}۱۵ \text{ نماد / بیت}.$$

(ب) احتمال $p(v_j)$ پیام معیّن v_j را می‌توان به عنوان نتیجهٔ این حقیقت که نمادهای تولید شده به طور آماری مستقلّند، به صورت حاصل‌ضرب احتمالهای نمادهای تکی نوشت، بنابراین برای احتمالهای $۹ = ۳^۲$ پیام ممکن از دو نماد نتیجه می‌شود که

$$p(v_۱) = p(u_۱, u_۱) = p(u_۱)p(u_۱) = ۰{,}۴۹,$$

$$p(v_۲) = p(u_۱, u_۲) = p(u_۱)p(u_۲) = ۰{,}۱۴,$$

$$p(v_۳) = p(u_۱, u_۳) = p(u_۱)p(u_۳) = ۰{,}۰۷,$$

$$p(v_۴) = p(u_۲, u_۱) = ۰{,}۱۴,$$

$$p(v_۵) = p(u_۲, u_۲) = ۰{,}۰۴,$$

$$p(v_۶) = p(u_۲, u_۳) = ۰{,}۰۲,$$

$$p(v_۷) = p(u_۳, u_۱) = ۰{,}۰۷,$$

$$p(v_۸) = p(u_۳, u_۲) = ۰{,}۰۲,$$

$$p(v_۹) = p(u_۳, u_۳) = ۰{,}۰۱.$$

(پ) با توجّه به (الف) داریم

$$H(V) = l\,H(U),$$

یعنی، یک پیام با طول l، اطلاعی به اندازهٔ l برابر یک پیام با طول ۱ را شامل می‌شود در صورتی که نمادهای متوالی به طور آماری مستقل باشند. از این رو

$$H(V) = ۲H(U) = ۲{,}۳۰ \text{ پیام / بیت}.$$

با توجّه به (ب) داریم

$$H(V) = -\sum_{j=۱}^{۹} p(v_j)\log p(v_j)$$

$$= -۰٫۴۹\log ۰٫۴۹ - ۰٫۱۴\log ۰٫۱۴ - ۰٫۰۷\log ۰٫۰۷$$

$$-۰٫۱۴\log ۰٫۱۴ - ۰٫۰۴\log ۰٫۰۴ - ۰٫۰۲\log ۰٫۰۲$$

$$-۰٫۰۷\log ۰٫۰۷ - ۰٫۰۲\log ۰٫۰۲ - ۰٫۰۱\log ۰٫۰۱$$

$$= ۲٫۳۰\ \text{بیت / پیام}.$$

از این رو هر دو نتیجه برابرند که موافق با قضیه می‌باشد.

(ت) حشو برابر است با

$$red = ۱ - \frac{H(U)}{\max\limits_{u} H(U)} = ۱ - \frac{۱٫۱۵}{\log ۳} = ۰٫۲۷.$$

(ث) برای اطلاع در ثانیه داریم

$$H_t(U) = \frac{۱}{t} H(U)\ \text{بیت / ثانیه},$$

که در آن t متوسط مدّت زمان یک نماد است.

$$t = ۰٫۷ \times ۰٫۰۰۱ + ۰٫۲ \times ۰٫۰۰۲ + ۰٫۱ \times ۰٫۰۰۳ = ۰٫۰۰۱۴\ \text{ثانیه}.$$

از این رو

$$H_t(U) = \frac{۱}{۰٫۰۰۱۴} ۱٫۱۵ = ۸۲۱٫۴۴\ \text{بیت / ثانیه}.$$

۲.۲ (الف) مقدار اطلاع بر نماد عبارت است از

$$H(U) = -\sum_{i=۱}^{۸} p(u_i)\log p(u_i)$$

$$= -\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۴}\log\frac{۱}{۴} - \frac{۱}{۸}\log\frac{۱}{۸} - \frac{۱}{۱۶}\log\frac{۱}{۱۶} - \frac{۱}{۳۲}\log\frac{۱}{۳۲}$$

$$-\frac{۱}{۶۴}\log\frac{۱}{۶۴} - ۲\times\frac{۱}{۱۲۸}\log\frac{۱}{۱۲۸} = ۱\frac{۶۳}{۶۴} = ۱٫۹۸\ \text{بیت}.$$

(ب) احتمال صفر را می‌توان به صورت زیر تعیین کرد:

$$p(۰) = \frac{\sum\limits_i p(u_i)c_{i۰}}{\sum\limits_i p(u_i)l_i},$$

که در آن c_{i0} تعداد صفرها در کدواژه و l_i تعداد نمادهایی است که این کدواژه را می‌سازند. از این نتیجه می‌شود

$$p(۰)=\left\{\frac{۱}{۲}۳+\frac{۱}{۴}۲+\frac{۱}{۸}۲+\frac{۱}{۱۶}۱+\frac{۱}{۳۲}۲+\frac{۱}{۶۴}۱+\frac{۱}{۱۲۸}۱\right\}\frac{۱}{۳}=۰٫۸.$$

بنابراین $p(۱)=۱-p(۰)=۰٫۲$.

(پ) برای کارایی داریم

$$\eta=\frac{H(U)}{L\log r}$$

که در آن L متوسط طول کدواژه‌ها می‌باشد. چون کدگذاری تعینی و یک به یک است، هیچ عدم‌حتمیتی معرّفی نشده است؛ بنابراین $H(U)$ مقدار اطلاع نمادهای منبع خروجی با اطلاع کدواژه‌ها یعنی بیت ۱٫۹۸ برابر است. با جای‌گذاری نتیجه می‌شود

$$\eta=\frac{۱٫۹۸}{۳\times۱}=۰٫۶۶.$$

(ت) با احتمالهای داده شده روش فانو و شانون به یک کد منجر می‌شود؛ نتیجه می‌شود:

نماد	احتمال	کد $(r=۲)$
$u_۱$	۱/۲ = ۶۴/۱۲۸	۰
$u_۲$	۱/۴ = ۳۲/۱۲۸	۱۰
$u_۳$	۱/۸ = ۱۶/۱۲۸	۱۱۰
$u_۴$	۱/۱۶ = ۸/۱۲۸	۱۱۱۰
$u_۵$	۱/۳۲ = ۴/۱۲۸	۱۱۱۱۰
$u_۶$	۱/۶۴ = ۲/۱۲۸	۱۱۱۱۱۰
$u_۷$	۱/۱۲۸ = ۱/۱۲۸	۱۱۱۱۱۱۰
$u_۸$	۱/۱۲۸ = ۱/۱۲۸	۱۱۱۱۱۱۱

(ث) برای تعیین کارایی ابتدا باید متوسط مدّت زمان کدواژه‌ها را محاسبه کرد. داریم

$$L=\frac{۱}{۲}۱+\frac{۱}{۴}۲+\frac{۱}{۸}۳+\frac{۱}{۱۶}۴+\frac{۱}{۳۲}۵+\frac{۱}{۶۴}۶+۲\times\frac{۱}{۱۲۸}۷=۱\frac{۶۳}{۶۴}=۱٫۹۸.$$

در این صورت کارایی برابر است با

$$\eta = \frac{۱٫۹۸}{۱٫۹۸ \times ۱} = ۱$$

۳.۲ (الف) کاربرد روش فانو برای یک کد سه‌سه‌ای با تقسیم پیوستهٔ نمادها بـه سـه گروه با احتمالهای تقریباً یکسان انجام می‌شود. در این روش کد زیر نتیجه می‌شود

نماد	احتمال	کد $(r = ۳)$
$u_۱$	۳/۸ = ۹/۲۴	۰
$u_۲$	۱/۶ = ۴/۲۴	۱۰
$u_۳$	۱/۸ = ۳/۲۴	۱۱
$u_۴$	۱/۸ = ۳/۲۴	۲۰
$u_۵$	۱/۸ = ۳/۲۴	۲۱
$u_۶$	۱/۱۲ = ۲/۲۴	۲۲

(ب) مقدار اطلاع برای هر نماد عبارت است از

$$H(U) = -\frac{۳}{۸}\log\frac{۳}{۸} - \frac{۱}{۶}\log\frac{۱}{۶} - ۳\times\frac{۱}{۸}\log\frac{۱}{۸} - \frac{۱}{۱۲}\log\frac{۱}{۱۲} = ۲٫۳۹ \text{ بیت}$$

به علاوه باید متوسط طول کدواژه را تعیین کرد. این برابر است با

$$L = \frac{۳}{۸}۱ + \frac{۱}{۶}۲ + ۳\times\frac{۱}{۸}۲ + \frac{۱}{۱۲}۲ = ۱٫۶۲۵.$$

در این حالت کارایی برابر است با

$$\eta = \frac{H(U)}{L\log r} = \frac{۲٫۳۹}{۱٫۶۲۵\log ۳} = ۰٫۹۳.$$

۴.۲ (الف) نمادها باید در حدّ امکان به چهار گروه تقسیم شوند. ایـن عمـل کـد I را می‌دهد:

نماد	احتمال	کد I	کد II
$u_۱$	۱/۴ = ۱۵/۶۰	a	a
$u_۲$	۱/۵ = ۱۲/۶۰	b	b
$u_۳$	۱/۶ = ۱۰/۶۰	ca	c
$u_۴$	۱/۶ = ۱۰/۶۰	cb	da
$u_۵$	۱/۱۲ = ۵/۶۰	da	db
$u_۶$	۱/۱۲ = ۵/۶۰	db	dc
$u_۷$	۱/۲۰ = ۳/۶۰	dc	dd

یادآوری: کد دیگر، یعنی کد II نتیجهٔ بهتری می‌دهد (یعنی یک کد با متوسط طول واژهٔ کمتر).

(ب) مقدار اطلاع برابر است با

$$H(U) = -\frac{1}{4}\log\frac{1}{4} - \frac{1}{5}\log\frac{1}{5} - 2\times\frac{1}{6}\log\frac{1}{6} - 2\times\frac{1}{12}\log\frac{1}{12} - \frac{1}{20}\log\frac{1}{20}$$

$= ۲٫۶۴$ بیت / نماد

متوسط طول کدواژهٔ کد I برابر است با

$$L = (\frac{1}{4}+\frac{1}{5})1 + (\frac{2}{6}+\frac{2}{12}+\frac{1}{20})2 = \frac{31}{20} = 1{,}55.$$

در این صورت کارایی این کد عبارت است از

$$\eta = \frac{H(U)}{L\log r} = \frac{2{,}64}{1{,}55\log 4} = 0{,}85.$$

با کد II کارایی برابر ۰٫۹۵ است.

۵.۲ (الف) با به کارگیری روش فانو بهترین کد به صورت زیر است

نماد	احتمال	کد $(r=3)$
u_1	$1/2 = 8/16$	a
u_2	$1/4 = 4/16$	b
u_3	$1/8 = 2/16$	ca
u_4	$1/16 = 1/16$	cb
u_5	$1/16 = 1/16$	cc

(ب) مقدار اطلاع در این حالت برابر است با

$$H(U) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - 2\times\frac{1}{16}\log\frac{1}{16}$$

$= \frac{15}{8} = ۱٫۸۷۵$ بیت / نماد.

متوسط طول کدواژه عبارت است از

$$L = (\frac{1}{2}+\frac{1}{4})1 + (\frac{1}{8}+2\times\frac{1}{16})2 = 1{,}25.$$

بنابراین کارایی این کد برابر است با

$$\eta = \frac{۱٫۸۷۵}{۱٫۲۵\log ۳} = ۰٫۹۵$$

۶.۲ (الف) روش فانو برای یافتن کدی مناسب انتخاب شده است:

نماد	احتمال	کد ($r = ۲$)
a	۱/۲ = ۳/۶	۰
b	۱/۳ = ۲/۶	۱۰
c	۱/۶ = ۱/۶	۱۱

(ب) مقدار اطلاع برابر است با

$$H(U) = -\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۳}\log\frac{۱}{۳} - \frac{۱}{۶}\log\frac{۱}{۶} = ۱٫۴۶ \text{ بیت / نماد.}$$

متوسط طول کدواژه برابر است با

$$L = \frac{۱}{۲}۱ + (\frac{۱}{۳} + \frac{۱}{۶})۲ = ۱٫۵.$$

بنابراین کارایی این کد به صورت زیر نتیجه می‌شود

$$\eta = \frac{۱٫۴۶}{۱٫۵\log ۲} = ۰٫۹۷.$$

(پ) کد را می‌توان با انتخاب پیوسته دو نماد با یکدیگر اصلاح نموده و نماد پیام جدیدی را تشکیل داد. چون این نمادها مستقلّند، احتمال نماد جدید برابر حاصل‌ضرب احتمالهای نمادهای اصلی است. اگر نمادهای جدید به ترتیب احتمال نزولی مرتّب شوند روش فانو کد زیر را می‌دهد (جوابهای بیشتری امکان‌پذیر است):

نماد	احتمال	کد ($r = ۲$)
aa	۱/۴ = ۹/۳۶	۰۰
ab	۱/۶ = ۶/۳۶	۰۱
ba	۱/۶ = ۶/۳۶	۱۰۰
bb	۱/۹ = ۴/۳۶	۱۰۱
ac	۱/۱۲ = ۳/۳۶	۱۱۰۰
ca	۱/۱۲ = ۳/۳۶	۱۱۰۱
bc	۱/۱۸ = ۲/۳۶	۱۱۱۰
cb	۱/۱۸ = ۲/۳۶	۱۱۱۱۰
cc	۱/۳۶ = ۱/۳۶	۱۱۱۱۱

(ت) مقدار اطلاع یک زوج نماد به اندازهٔ دو برابر یک نماد است زیرا نمادها از یکدیگر

مستقلّند، از این رو: بیت $H(V)=۲\times ۱٫۴۶=۲٫۹۲$.

اکنون متوسط طول کدواژه‌ها برابر می‌شوند با

$$L=(\frac{۱}{۴}+\frac{۱}{۶})۲+(\frac{۱}{۶}+\frac{۱}{۹})۳+(\frac{۱}{۱۲}+\frac{۱}{۱۲}+\frac{۱}{۱۸})۴+(\frac{۱}{۱۸}+\frac{۱}{۳۶})۵$$

بیت برای هر زوج نماد (یا ۱٫۴۹ برای هر نماد اصلی) $=۲٫۹۷$

که توسط آن داریم

$$\eta=\frac{۲٫۹۲}{۲٫۹۷\log ۲}=۰٫۹۸.$$

از این رو با اختیار نمادها با یکدیگر کارایی مقداری بهبود می‌یابد.

۷.۲ (الف) روش فانو کد زیر را می‌دهد

نماد	احتمال	کد $(r=۳)$
$u_۱$	۱/۴ = ۸/۳۲	a
$u_۲$	۱/۴ = ۸/۳۲	ba
$u_۳$	۱/۸ = ۴/۳۲	bb
$u_۴$	۱/۸ = ۴/۳۲	ca
$u_۵$	۱/۱۶ = ۲/۳۲	cba
$u_۶$	۱/۱۶ = ۲/۳۲	cbb
$u_۷$	۱/۱۶ = ۲/۳۲	cca
$u_۸$	۱/۳۲ = ۱/۳۲	ccb
$u_۹$	۱/۳۲ = ۱/۳۲	ccc

روش دیگر روش هافمن است؛ این روش کد زیر را می‌دهد

نماد	احتمال				کد $(r=۳)$
$u_۱$	۸/۳۲	۸/۳۲	۸/۳۲	۱۶/۳۲ (a)	b
$u_۲$	۸/۳۲	۸/۳۲	۸/۳۲	۸/۳۲ (b)	c
$u_۳$	۴/۳۲	۴/۳۲	۸/۳۲ (a)	۸/۳۲ (c)	ab
$u_۴$	۴/۳۲	۴/۳۲	۴/۳۲ (b)		ac
$u_۵$	۲/۳۲	۴/۳۲ (a)	۴/۳۲ (c)		aab
$u_۶$	۲/۳۲	۲/۳۲ (b)			aac
$u_۷$	۲/۳۲ (a)	۲/۳۲ (c)			$aaaa$
$u_۸$	۱/۳۲ (b)				$aaab$
$u_۹$	۱/۳۲ (c)				$aaac$

مقدار اطلاع برای هر نماد برابر است با

$$H(U) = -۲\times\frac{۱}{۴}\log\frac{۱}{۴} - ۲\times\frac{۱}{۸}\log\frac{۱}{۸} - ۳\times\frac{۱}{۱۶}\log\frac{۱}{۱۶} - ۲\times\frac{۱}{۳۲}\log\frac{۱}{۳۲}$$

$$= ۱ + \frac{۳}{۴} + \frac{۳}{۴} + \frac{۵}{۱۶} = ۲{,}۸۲ \text{ نماد / بیت.}$$

متوسط طول کد برای روش فانو برابر است با

$$L = ۱\times\frac{۱}{۴} + ۲(\frac{۱}{۴} + \frac{۱}{۸} + \frac{۱}{۸}) + ۳(\frac{۱}{۱۶} + \frac{۱}{۱۶} + \frac{۱}{۱۶} + \frac{۱}{۳۲} + \frac{۱}{۳۲}) = ۲,$$

و برای روش هافمن عبارت است از

$$L = ۱(\frac{۱}{۴} + \frac{۱}{۴}) + ۲(\frac{۱}{۸} + \frac{۱}{۸}) + ۳(\frac{۱}{۱۶} + \frac{۱}{۱۶}) + ۴(\frac{۱}{۱۶} + \frac{۱}{۳۲} + \frac{۱}{۳۲}) = \frac{۱۵}{۸} = ۱{,}۸۸.$$

بنابراین کارایی برطبق روش فانو برابر می‌شود با $\eta_F = \frac{۲{,}۸۲}{۲\log ۳} = ۰{,}۸۹$.

و برطبق روش هافمن عبارت است از $\eta_H = \frac{۲{,}۸۲}{۱{,}۸۸\log ۳} = ۰{,}۹۵$.

در نتیجه روش هافمن به کد کاراتری منجر می‌شود.

(ب) اگر کدنماد c نبایستی با c دنبال شود نمی‌توان اجازه داد که

- ترکیب cc در یک کدواژه ظاهر شود.
- یک کدواژه با c ختم شود، اگر بیش از یک کدواژه با c شروع شود.

یک کد ممکن به صورت زیر است

نماد	احتمال	کد $(r = ۳)$
$u_۱$	۱/۴ = ۸/۳۲	a
$u_۲$	۱/۴ = ۸/۳۲	ba
$u_۳$	۱/۸ = ۴/۳۲	bb
$u_۴$	۱/۸ = ۴/۳۲	caa
$u_۵$	۱/۱۶ = ۲/۳۲	cab
$u_۶$	۱/۱۶ = ۲/۳۲	cba
$u_۷$	۱/۱۶ = ۲/۳۲	cbb
$u_۸$	۱/۳۲ = ۱/۳۲	$abca$
$u_۹$	۱/۳۲ = ۱/۳۲	$cbcb$

متوسط طول برای این کد برابر است با

$$L = \frac{۱}{۴} + ۲ \times (\frac{۱}{۴} + \frac{۱}{۸}) + ۳ \times (\frac{۱}{۸} + \frac{۳}{۱۶}) + ۴ \times (\frac{۱}{۳۲} + \frac{۱}{۳۲}) = ۲{,}۸۲.$$

بنابراین کارایی برابر است با

$$\eta = \frac{۲{,}۸۲}{۲{,}۸۲ \log ۳} = ۰{,}۸۱.$$

۸.۲ (الف)، (ب) با ترکیب دو پاسخ ۹ امکان به وجود می‌آید. پس از ترتیب مجدّد به صورت احتمال نزولی کدهای زیر با روش فانو به دست می‌آید:

نماد	احتمال	$(r = ۲)$ کد (الف)	$(r = ۳)$ کد (ب)
بلی و بلی	۰٫۳۰	۰۰	a
بلی و خیر	۰٫۲۴	۰۱	ba
خیر و بلی	۰٫۱۰	۱۰۰	bb
بدون نظر و بلی	۰٫۱۰	۱۰۱	ca
خیر و خیر	۰٫۰۸	۱۱۰۰	cba
بدون نظر و خیر	۰٫۰۸	۱۱۰۱	cbb
بلی و بدون نظر	۰٫۰۶	۱۱۱۰	cca
خیر و بدون نظر	۰٫۰۲	۱۱۱۱۰	ccb
بدون نظر و بدون نظر	۰٫۰۲	۱۱۱۱۱	ccc

(پ) مقدار اطلاع پیام با مجموع مقدار اطلاع نسبت به سؤال ۱ و سؤال ۲ برابر است. این نتیجه می‌دهد

$$H(V) = -۰{,}۵ \log ۰{,}۵ - ۰{,}۴ \log ۰{,}۴ - ۰{,}۱ \log ۰{,}۱$$

$$- ۰{,}۶ \log ۰{,}۶ - ۰{,}۲ \log ۰{,}۲ - ۰{,}۲ \log ۰{,}۲ = ۲{,}۷۳ \text{ بیت}.$$

برای کد دودویی در (الف) متوسط طول برابر است با

$$L = ۲ \times ۰{,}۵۴ + ۳ \times ۰{,}۲۰ + ۴ \times ۰{,}۲۲ + ۵ \times ۰{,}۰۴ = ۲{,}۷۶,$$

بنابراین کارایی برابر می‌شود با:

$$\eta_a = \frac{۲{,}۷۳}{۲{,}۷۶ \log ۲} = ۰{,}۹۹.$$

برای کد سه‌سه‌ای در (ب) متوسط طول برابر است با

$$L = ۱ \times ۰٫۳۰ + ۲ \times ۰٫۴۴ + ۳ \times ۰٫۲۶ = ۱٫۹۶,$$

بنابراین در این جا کارایی برابر است با

$$\eta_b = \frac{۲٫۷۳}{۱٫۹۶ \log ۳} = ۰٫۸۸.$$

بنابراین براساس کارایی کد دودویی ترجیح داده می‌شود.

۹.۲ نمادهای u_1، ...، u_4 با احتمالهای مربوط به خودشان در جدول داده شده‌اند.

نماد	احتمال
u_1	۰٫۱
u_2	۰٫۳
u_3	۰٫۲
u_4	۰٫۴

یک کد برطبق روش ژیلبرت - مور با تعیین طول کدواژه برای هر نماد به دست آورده و سپس یک سری صعودی α_i که در این صورت وابسته به یک سری دودویی است می‌سازیم. برای نماد u_1 داریم

$$۲^{-۴} < p(u_1) = \frac{۱}{۱۰} < ۲^{-۳},$$

بنابراین نمادهای (دودویی) با طول $l_1 = ۵$ لازم است. مقدار $\alpha_1 = \frac{۱}{۲} p(u_1)$ برابر می‌شود با $\alpha_1 = ۰٫۰۵$. اگر این عدد به صورت دودویی نوشته و پس از ۵ عدد قطع شود در این صورت عدد ۰۰۰۰۱ به عنوان کدواژه به دست می‌آید. سه کدواژهٔ باقی‌مانده را به همین روش تعیین می‌کنیم. جدول زیر نتیجهٔ کد را می‌دهد:

	احتمال	l_i	α_i	کد
u_1	۰٫۱	۵	۰٫۰۵	۰۰۰۰۱
u_2	۰٫۳	۳	۰٫۲۵	۰۱۰
u_3	۰٫۲	۴	۰٫۵۰	۱۰۰۰
u_4	۰٫۴	۳	۰٫۸۰	۱۱۰

برای به دست آوردن کارایی این کد ابتدا باید محتوای اطلاع منبع را تعیین کرد، که به صورت زیر داده می‌شود:

$$H(U) = -\sum_{i=1}^{4} p(u_i)\log p(u_i)$$
$$= -0.1\log 0.1 - 0.3\log 0.3 - 0.2\log 0.2 - 0.4\log 0.4$$
$$= 1.85 \text{ بیت / نماد}.$$

متوسط طول کدواژه برابر است با:

$$L = 0.1\times 5 + 0.3\times 3 + 0.2\times 4 + 0.4\times 3 = 3.4,$$

بنابراین کارایی برابر است با:

$$\eta_G = \frac{1.85}{3.4\times 1} = 0.54.$$

پس از ترتیب مجدّد نمادها برطبق احتمال صعودی، یک کد برطبق روش شانون به صورت زیر به دست می‌آید:

	احتمال	P_i	l_i	کد $(r=2)$
u_1	۰٫۴	۰	۲	۰۰
u_2	۰٫۳	۰٫۴	۲	۰۱
u_3	۰٫۲	۰٫۷	۳	۱۰۱
u_4	۰٫۱	۰٫۹	۴	۱۱۱۰

اکنون متوسط طول کدواژه عبارت است از

$$L_S = 0.4\times 2 + 0.3\times 2 + 0.2\times 3 + 0.1\times 4 = 2.4,$$

که کارایی زیر را می‌دهد

$$\eta_S = \frac{1.85}{2.4\times 1} = 0.77.$$

در پایان، کد برطبق روش هافمن با انتخاب پیوسته دو نماد با کمترین احتمال رخداد با همدیگر به دست می‌آید. با این روش کد زیر نتیجه می‌شود:

نماد	احتمال			کد $(r=2)$
u_1	۰٫۴	۰٫۴	۰٫۶ (۰)	۱
u_2	۰٫۳	۰٫۳ (۰)	۰٫۴ (۱)	۰۰
u_3	۰٫۲ (۰)	۰٫۳ (۱)		۰۱۰
u_4	۰٫۱ (۱)			۰۱۱

متوسط طول کد برطبق روش هافمن برابر است با

$$L_H = ۰٫۴ \times ۱ + ۰٫۳ \times ۲ + ۰٫۲ \times ۳ + ۰٫۱ \times ۳ = ۱٫۹,$$

بنابراین کارایی عبارت است از

$$\eta_H = \frac{۱٫۸۵}{۱٫۹ \times ۱} = ۰٫۹۷.$$

نتیجه می‌شود که روش هافمن بهترین کارایی را می‌دهد و روش ژیلبرت-مور بدترین بین سه تا می‌باشد؛ با وجود این، این کد با یک شرط اضافی مواجه می‌شود، یعنی این که ترتیب نمادها یکسان باقی می‌ماند.

۲.۱۰ (الف) برای هر دو کدهای هافمن و فانو لازم است که نمادها را به ترتیب احتمال نزولی رخ‌داد مرتّب کرد. برای مثال یک کد هافمن دودویی به صورت زیر است:

نماد	احتمال						کد $(r=2)$
u_1	۱/۳	۱/۳	۱/۳	۱/۳	۱/۳	۲/۳ (۰)	۱
u_2	۱/۳	۱/۳	۱/۳	۱/۳	۱/۳ (۰)	۱/۳ (۱)	۰۰
u_3	۱/۹	۱/۹	۱/۹	۲/۹ (۰)	۱/۳ (۱)		۰۱۱
u_4	۱/۹	۱/۹	۱/۹ (۰)	۱/۹ (۱)			۰۱۰۰
u_5	۱/۲۷	۲/۲۷ (۰)	۱/۹ (۱)				۰۱۰۱۱
u_6	۱/۲۷ (۰)	۱/۲۷ (۱)					۰۱۰۱۰۰
u_7	۱/۲۷ (۱)						۰۱۰۱۰۱

برای تعیین کارایی، ابتدا مقدار اطلاع را محاسبه می‌کنیم:

$$H(U) = -۲ \times \frac{۱}{۳}\log\frac{۱}{۳} - ۲ \times \frac{۱}{۹}\log\frac{۱}{۹} - ۳ \times \frac{۱}{۲۷}\log\frac{۱}{۲۷} = ۲٫۲۹ \text{ نماد / بیت}.$$

متوسط طول برابر است با:

$$L_H = \frac{۱}{۳} \times (۱+۲) + \frac{۱}{۹} \times (۳+۴) + \frac{۱}{۲۷} \times (۵+۶+۶) = ۶۵/۲۷ = ۲{,}۴۱.$$

بنابراین کارایی برابر است با:

$$\eta_H = \frac{۲{,}۲۹}{۲{,}۴۱ \times ۱} = ۰{,}۹۵.$$

(ب) برای تعیین یک کد سه‌سه‌ای فانو نمادها را همواره به سه گروه تقریباً هم‌احتمال تقسیم می‌کنیم. کدنمادها را با a، b و c نشان می‌دهیم.

نماد	احتمال	کد $(r=۳)$
$u_۱$	۱/۳	a
$u_۲$	۱/۳	b
$u_۳$	۱/۹	ca
$u_۴$	۱/۹	cb
$u_۵$	۱/۲۷	cca
$u_۶$	۱/۲۷	ccb
$u_۷$	۱/۲۷	ccc

متوسط طول این کد برابر است با

$$L_F = \frac{۱}{۳} \times (۱+۱) + \frac{۱}{۹} \times (۲+۲) + \frac{۱}{۲۷} \times (۳+۳+۳) = \frac{۱۳}{۹},$$

که کارایی آن عبارت است از

$$\eta_F = \frac{۲{,}۲۹}{۱{,}۴۴ \times ۱{,}۵۸} = ۱{,}۰۰.$$

(پ) کد سه‌سه‌ای فانو براساس کارایی برتر می‌باشد. با وجود این، اگر هزینه برای هر کدنماد مورد توجّه قرار داده شود در این صورت معیار کارایی ماکسیمم به معیار می‌نیمم متوسط هزینه تبدیل می‌شود.

$$Costs_H = (۶۵/۲۷) \times ۱{,}۸۰ = ۴{,}۳۳ \text{ (پوند)}.$$

هزینه کد فانو عبارت است از

$$Costs_F = (۱۳/۹) \times ۲{,}۷۰ = ۳{,}۹۰ \text{ (پوند)}.$$

بنابراین در این حالت نیز کد فانو از مزیّت برتری بهره‌مند می‌گردد.

۱۱.۲ (الف) در طرف رمزگشایی یک سری اعداد دریافت می‌شود که می‌توان بدون انتظار برای نماد بعدی مستقیماً به دنباله‌ای از نمادهای دودویی برگرداند. از این رو کد مستقیماً قابل کدگشایی است. کد نیز قابل کدگشایی فوری است، زیرا یک رابطهٔ یک به یک صریحی بین نمادهای منبع دودویی و کدنمادها وجود دارد حتّی اگر آنها در دنباله‌ای از نمادها رخ دهند.

(ب) اگر احتمالهای کدنمادها محاسبه شده باشند محتوای اطلاع برای هر کدنماد را می‌توان تعیین کرد. با احتمالهای داده شدهٔ نمادهای دودویی این احتمالها را می‌توان به صورت زیر تعیین کرد؛ مثلاً

$$p(۳) = p(۰)^۳ \, p(۱) = ۰٫۸^۳ \times ۰٫۲ = ۰٫۱۰۲۴.$$

از این رو احتمالهای زیر به دست می‌آیند:

کدنماد	احتمال
۰	۰٫۲
۱	۰٫۱۶
۲	۰٫۱۲۸
۳	۰٫۱۰۲۴
۴	۰٫۴۰۹۶

بنابراین محتوای اطلاع برای هر کدنماد برابر است با

$$H(U) = -۰٫۲ \log ۰٫۲ - ۰٫۱۶ \log ۰٫۱۶ - ۰٫۱۲۸ \log ۰٫۱۲۸$$

$$- ۰٫۱۰۲۴ \log ۰٫۱۰۲۴ - ۰٫۴۰۹۶ \log ۰٫۴۰۹۶ = ۲٫۱۳ \text{ بیت / نماد}$$

(پ) کارایی به صورت زیر است

$$\eta = \frac{H(U)}{L \log r} = \frac{۲٫۱۳}{۱ \times \log ۵} = ۰٫۹۲.$$

۱۲.۲ (الف) روش شانون احتمال تجمعی P_k را به کار می‌برد. طول کدواژه‌ها از نابرابری زیر به دست می‌آیند:

$$\log\frac{۱}{p_k} \le l_k \le \log\frac{۱}{p_k} + ۱.$$

نتیجهٔ نهایی به صورت زیر است

نماد	p_i	P_i	l_i	کد $(r = ۲)$
$u_۱$	۰٫۳۲	۰	۲	۰۰
$u_۲$	۰٫۲۴	۰٫۳۲	۳	۰۱۱
$u_۳$	۰٫۲۰	۰٫۵۶	۳	۱۰۰
$u_۴$	۰٫۰۹	۰٫۷۶	۴	۱۱۰۰
$u_۵$	۰٫۰۵	۰٫۸۵	۵	۱۱۰۱۱
$u_۶$	۰٫۰۴	۰٫۹۰	۵	۱۱۱۰۱
$u_۷$	۰٫۰۴	۰٫۹۴	۵	۱۱۱۱۰
$u_۸$	۰٫۰۲	۰٫۹۸	۶	۱۱۱۱۱۱

(ب) برای سه کدنماد تعداد نمادهای پیام بایستی برابر با ۲k+۳ باشد تا به یک کد هافمن بهینه برسیم. بنابراین باید یک نماد ساختگی که احتمال صفر دارد اضافه کنیم؛ لذا به دست می‌آوریم

نماد	احتمال				کد $(r = ۳)$
$u_۱$	۰٫۳۲	۰٫۳۲	۰٫۳۲	۰٫۴۴ (a)	b
$u_۲$	۰٫۲۴	۰٫۲۴	۰٫۲۴	۰٫۳۲ (b)	c
$u_۳$	۰٫۲۰	۰٫۲۰	۰٫۲۰ (a)	۰٫۲۴ (c)	aa
$u_۴$	۰٫۰۹	۰٫۰۹	۰٫۱۵ (b)		ac
$u_۵$	۰٫۰۵	۰٫۰۶ (a)	۰٫۰۹ (c)		abb
$u_۶$	۰٫۰۴	۰٫۰۵ (b)			abc
$u_۷$	۰٫۰۴ (a)	۰٫۰۴ (c)			$abaa$
$u_۸$	۰٫۰۲ (b)				$abab$
$u_۹$	۰٫۰۰ (c)				----

(پ) روش فانو برای کد چهار مقداری براساس فرایند تقسیم نمادها به چهار گروه با احتمالهای تقریباً برابر برای هر گروه می‌باشد.

نماد	احتمال	کد $(r=4)$
u_1	۰٫۳۲	a
u_2	۰٫۲۴	b
u_3	۰٫۲۰	c
u_4	۰٫۰۹	da
u_5	۰٫۰۵	db
u_6	۰٫۰۴	dc
u_7	۰٫۰۴	dda
u_8	۰٫۰۲	ddb

(ت) لازم نیست برای مقایسه سه کد کارایی کامل را تعیین کنیم. با مقایسه $L\log r$ می‌توان آن را انجام داد. از این رو داریم

کد	L	$\log r$	$L\log r$
I	۳٫۰۹	۱	۳٫۰۹
II	۱٫۶۵	۱٫۵۸	۲٫۶۱
III	۱٫۳۰	۲	۲٫۶۰

براساس این جدول می‌توان نتیجه گرفت که کد فانو بهترین است، ولو این که اختلاف کمی نسبت به کد هافمن دارد.

(ث) با کد دودویی یافته شده در (الف) برای احتمال صفر نتیجه می‌شود که

$$p(0)=\frac{1}{L}\sum_{i=1}^{8}p(u_i)c_i(0),$$

که در آن $c_i(0)$ تعدا صفرها در کدواژه برای نماد u_i است. نتیجه می‌شود که

$$p(0)=\frac{1}{3{,}09}[0{,}32\times 2+0{,}24\times 1+0{,}20\times 2+0{,}09\times 2+0{,}05\times 1$$
$$+0{,}04\times 1+0{,}04\times 1+0{,}02\times 0]=0{,}5,$$

و از این رو $p(1)=1-p(0)=0{,}5$.

بنابراین محتوای اطلاع برای هر کدنماد دودویی برابر است با

$$H(U)=-0{,}5\log 0{,}5-0{,}5\log 0{,}5=1\ \text{بیت}.$$

۳

منبع اطلاع گسسته باحافظه

۱.۳ فرایندهای مارکوف

در فصل دوم فرض شد که منابع اطلاع حافظه ندارند؛ این بدین معنی است که نمادهای متوالی در یک پیام تولید شده توسط منبع به طور آماری مستقلند. در بسیاری از کاربردهای عملی این چنین نیست و احتمال وقوع یک نماد در یک پیام بستگی به تعداد متناهی از نمادهای قبلی خواهد داشت. در این حالت می‌توان از منبع اطلاع با حافظه نام برد. دنبالهٔ تولید شده توسط چنین منبعی را می‌توان به عنوان به اصطلاح ***زنجیر مارکوف*** در نظر گرفت. قبل از این‌که منبع اطلاع با حافظه را بررسی کنیم، برخی از ویژگیهای زنجیرهای مارکوف را به تفصیل مورد توجّه قرار خواهیم داد.

دنباله‌ای از متغیّرهای تصادفی گسسته $\mathbf{u}_1$، $\mathbf{u}_2$، ... ، $\mathbf{u}_{n-1}$ را در نظر می‌گیریم؛ برای مثال، اینها می‌توانند نمادهایی از یک منبع اطلاع گسسته یا نمونه‌های کمّی شدهٔ یک سیگنال مستقل باشند. نخست $\mathbf{u}_1$ به تمام نمادهایی که می‌توانند نماد اوّل باشند، $\mathbf{u}_2$ به تمام نمادهایی که می‌توانند نماد دوم باشند و الی‌آخر نسبت داده می‌شود. در حالت منابع بی‌حافظه تمام نمادهای پیامهای تولید شده می‌توانند تمام نمادهای الفبای منبع باشند. در حالت منابع باحافظه اتفاق می‌افتد که به علّت رخداد اوّلین نماد، دومین نماد تنها می‌تواند عضوی از زیرمجموعهٔ محدود شده‌ای از الفبای منبع باشد. اکنون توزیع احتمال $\mathbf{u}_n$، برای یک ***زنجیر مارکوف مرتبهٔ*** k، که بستگی به k مقدار متغیّرهای تصادفی قبلی $\mathbf{u}_{n-k}$، $\mathbf{u}_{n-(k-1)}$، ... ، $\mathbf{u}_{n-1}$ دارد مورد توجّه است، که در آن k به قسمی انتخاب می‌شود

که

(الف) مقادیر متغیّرهای تصادفی قبل از $\mathbf{u}_{n-k}$ هیچ اثری بر توزیع احتمال $\mathbf{u}_n$ نداشته باشند.
(ب) k می‌نیمم مقداری است که برای آن (الف) معتبر است.

در این صورت احتمال شرطی مقدار u_n از متغیّر تصادفی $\mathbf{u}_n$ به فرض معلوم بودن تمام مقادیر قبلی u_{n-1}، u_{n-2}، ... برابر است با $P(u_n|u_{n-k},u_{n-k+1},\dots,u_{n-1})$. چون ترکیب مقادیر $S_i=(u_{n-k},u_{n-k+1},\dots,u_{n-1})$ قسمت شرطی توزیع احتمال شرطی $\mathbf{u}_n$ را تشکیل می‌دهد، S_i را ***حالت زنجیر مارکوف*** می‌نامند. اگر مقدار u_n رخ دهد، آن‌گاه زنجیر مارکوف به حالت جدید S_j منتقل می‌شود:

$$S_j=(u_{n-(k-1)},u_{n-(k-2)},\dots,u_n).$$

در این صورت گوییم حالت S_i به حالت S_j ***انتقال می‌یابد***. علاوه بر این با معلوم بودن توزیع احتمال شرطی $\mathbf{u}_n$، همچنین می‌توان با دادن ماتریس احتمالهای انتقال برای حالتهای مختلف زنجیر مارکوف را مشخّص کرد. احتمال این را که زنجیر مارکوف از حالت S_i به حالت S_j منتقل شود با $P(j|i)$ نمایش می‌دهیم.

در بخش (۱.۲) به حالتی توجّه کردیم که در آن l نماد تولید شده توسط یک منبع اطلاع دائماً به عنوان پیام جدیدی در نظر گرفته می‌شد. در این صورت در واقع با یک منبع اطلاع جدید که دارای الفبای V است سروکار داریم که در آن نماد ترکیبی از نمادهای الفبای اصلی U است. اگر ترکیب S_i از k مقدار از متغیّرهای تصادفی در یک روش مشابه به عنوان مقدار S_i از یک متغیّر تصادفی جدید در نظر گرفته شود در این صورت احتمال حالت S_j تنها به حالت S_i بستگی دارد. به این طریق یک زنجیر مارکوف از مرتبهٔ قبلی k برحسب نمادها به یک زنجیر مارکوف از مرتبهٔ یک برحسب حالتها تبدیل می‌گردد. گرچه مارکوف ریاضی‌دان روسی زنجیر را وقتی به کار برده است که در آن مرتبهٔ k بایستی بزرگتر از ۱ باشد، در متون ریاضی جاری صرفاً زنجیر مارکوف مرتبهٔ $k=1$ به کار برده می‌شود. با وجود این، در نظریهٔ اطلاع دقیقاً زنجیرهای مارکوف مرتبهٔ بالاتر مهم هستند، زیرا در این صورت گاهی اوقات توصیف منبع اطلاع ساده‌تر می‌شود.

در این‌جا اگر فرض کنیم که هر متغیّر تصادفی $\mathbf{u}_i$ دارای m برآمد ممکن است، یک زنجیر مارکوف مرتبهٔ k را می‌توان در m^k حالت مختلف یافت زیرا هر حالتی با دنباله‌ای از k نماد که هر یک از m امکان انتخاب شده‌اند تعیین می‌شود، چون بعد از هر حالت، انتخابی از m امکان می‌توان داشت در این صورت m^{k+1} انتقال قابل تصوّر با تعدادی مساوی

احتمال انتقال وجود دارد. از هر گروه m تایی از احتمال انتقال، $m-1$ احتمال را می‌توان آزادانه انتخاب کرد. در این صورت احتمال باقی‌مانده ثابت است، چون مجموع احتمالهای انتقال برابر یک است. بنابراین m^{k+1} احتمال انتقال با انتخاب آزادانه $m^{k+1} - m^k$ احتمال انتقال تعیین می‌گردد.

می‌توان حالتهای این چنین تعریف شده را با احتمالهای انتقالشان در یک ***نمودار حالت*** رسم کرد. می‌توان دنباله‌ای از نمادهای به طور آماری مستقل را به عنوان ساده‌ترین حالت در نظر گرفت. در این صورت چون k برابر صفر است می‌توان از یک زنجیر مارکوف مرتبهٔ صفر صحبت کرد. این در واقع یک تباهیدگی زنجیر مارکوف است که برطبق تعریف براساس وجود احتمالهای انتقال است. چنین زنجیری فقط یک حالت S خواهد داشت؛ پس از هر «انتقال» زنجیر به همان حالت باز می‌گردد. تعداد انتقالهای ممکن برابر تعداد نمادهایی است که می‌توان انتخاب کرد. برای سه نماد (a, b, c) این زنجیر را می‌توان همان‌طور که در شکل (۱.۳) نشان داده شده است، نمایش داد.

برای یک زنجیر مارکوف از مرتبهٔ ۱، تعداد حالتها برابر تعداد نمادهاست. اگر این تعداد ۳ باشد، یعنی a، b و c، در این صورت تعداد انتقالها $9 = 3^2$ می‌باشند، یعنی $a \to a$، $a \to b$، $a \to c$، $b \to a$ و الی‌آخر. چنین زنجیر مارکوفی را می‌توان همانند شکل (۲.۳) نشان داد.

در این مثال حالتهای S_1، S_2 و S_3 را می‌توان با a، b و c نشان داد. مسلماً، برای هر حالت S_i، $i = 1,2,3$، داریم

$$P(S_1 | S_i) + P(S_2 | S_i) + P(S_3 | S_i) = 1. \qquad (۱.۳)$$

احتمالهای سه حالت S_1، S_2 و S_3 را می‌توان از احتمالهای انتقال به دست آورد

$$P(S_i) = P(S_1) \cdot P(S_i | S_1) + P(S_2) \cdot P(S_i | S_2) + P(S_3) \cdot P(S_i | S_3), \qquad (۲.۳)$$

برای $i = 1,2,3$.

راجع به این عبارت باید به مطلب زیر توجّه کرد. در بخش (۲.۱) (معادلهٔ (۱۴.۱) را ببینید) قضیهٔ بیز را بررسی کردیم و نشان دادیم که چگونه می‌توان احتمال شرطی $p(x_i | y_j)$ را براساس احتمال شرطی داده شده $q(y_j | x_i)$ حساب کرد: با وجود این، این قضیه را نمی‌توان برای احتمالهای انتقال در زنجیر مارکوف به کار برد. دلیل آن این است که در واقع قضیهٔ بیز فرض می‌کند که برای i و j ها

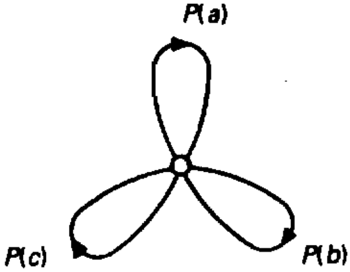
$$p(x_i, y_j) = p(y_j, x_i).$$

ولی برای زنجیر مارکوف معمولاً چنین موردی وجود ندارد. در عوض داریم

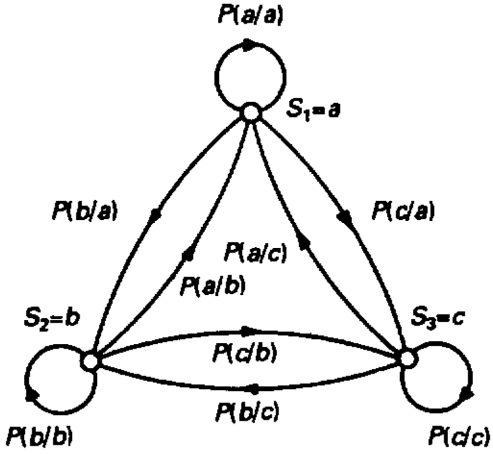
$$P(S_i,S_j) \neq P(S_j,S_i). \qquad (۳.۳)$$

دلیلش این است که موقعیت زمان برای زنجیر مارکوف و به تبع آن برای مرتب کردن نمادها نقش اساسی بازی می‌کند. اگر عبارت نوشته‌ای را در نظر بگیریم (که در واقع می‌توان به عنوان فرایند مارکوفی در نظر گرفت که وجود حروف توسط حروف قبلی تعیین می‌شوند) در این صورت واضح است که احتمال زوج حرف (q,u) برابر احتمال زوج حرف (u,q) نخواهد بود.

اهمیت زنجیرهای مارکوف به دلیل این حقیقت است که k، تعداد متغیّرهای تصادفی، که مقادیر آن احتمال انتقال به متغیّر تصادفی بعدی u_n راتعیین می‌کند، متناهی است، به قسمی که باید گذشته را ***برای همیشه*** نادیده گرفت.



شکل ۱.۳- نمودار حالت برای یک زنجیر مارکوف از مرتبهٔ صفر



شکل ۲.۳- نمودار حالت برای یک زنجیر مارکوف از مرتبهٔ ۱

مثال ۱.۳

یک منبع اطلاعی را که زنجیر مارکوف تولید می‌کند در نظر بگیرید. الفبای منبع $U = \{۰,۱\}$ است. به علاوه احتمالهای انتقال زیر داده شده‌اند

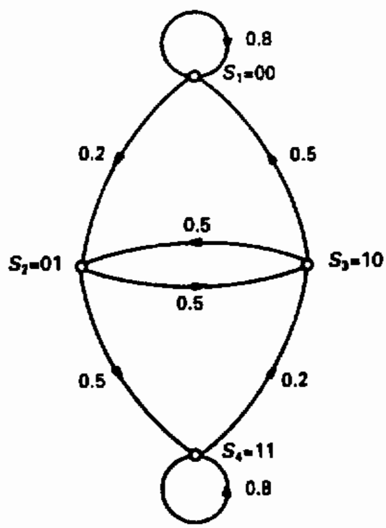
$$P(۰|۰۰) = P(۱|۱۱) = ۰٫۸,$$

$$P(۱|۰۰) = P(۰|۱۱) = ۰٫۲,$$

$$P(۰|۰۱) = P(۰|۱۰) = P(۱|۰۱) = P(۱|۱۰) = ۰٫۵.$$

در این‌جا می‌توان نتیجه گرفت که با یک زنجیر مارکوف از مرتبهٔ ۲ سروکار داریم. رخ‌داد هر نماد با رخ‌داد دو نماد قبلی تعیین می‌شود. در این صورت چهار حالت ۰۰، ۰۱، ۱۰ و ۱۱ وجود دارد. نمودار حالت در شکل (۳.۳) رسم شده است. از شکل می‌توان نتیجه گرفت که مستقیماً نمی‌توان از هر حالتی به حالت دیگر رسید. از این رو می‌توان از $S_۱$ به $S_۲$ رسید ولی نمی‌توان به $S_۳$ و $S_۴$ رفت. گاهی اوقات می‌توان از یک حالت به حالت دیگر رفت ولی نمی‌توان برگشت؛ می‌توانیم از $S_۱$ به $S_۲$ برویم ولی نمی‌توانیم مستقیماً از $S_۲$ به $S_۱$ برویم. احتمالهای حاشیه‌ای حالتهای $S_۱$ تا $S_۴$ را می‌توان براساس معادلهٔ (۲.۳) محاسبه کرد.

$$\begin{aligned} P(S_۱) &= P(S_۱)\,.\,P(S_۱|S_۱) + P(S_۲)\,.\,P(S_۱|S_۲) + P(S_۳)\,.\,P(S_۱|S_۳) + P(S_۴)\,.\,P(S_۱|S_۴) \\ &= P(S_۱)\times ۰٫۸ \quad + P(S_۲)\times ۰ \quad + P(S_۳)\times ۰٫۵ \quad + P(S_۴)\times ۰ \\ &= ۰٫۸P(S_۱) + ۰٫۵P(S_۳) \end{aligned}$$



شکل ۳.۳- نمودار حالت برای مثال (۱.۳)

به همین روش می‌توان به دست آورد که

$$P(S_2) = P(S_1)\times 0.2 + P(S_2)\times 0 + P(S_3)\times 0.5 + P(S_4)\times 0$$
$$= 0.2P(S_1) + 0.5P(S_3)$$
$$P(S_3) = P(S_1)\times 0 + P(S_2)\times 0.5 + P(S_3)\times 0 + P(S_4)\times 0.2$$
$$= 0.5P(S_2) + 0.2P(S_4)$$
$$P(S_4) = P(S_1)\times 0 + P(S_2)\times 0.5 + P(S_3)\times 0 + P(S_4)\times 0.8$$
$$= 0.5P(S_2) + 0.8P(S_4).$$

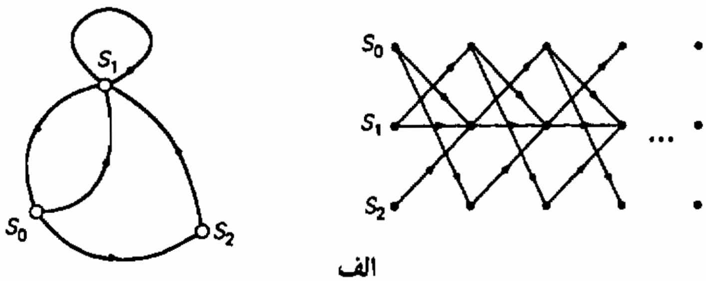
با حلّ این چهار معادله با چهار مجهول به دست می‌آوریم

$$P(S_1) = P(S_4) = \frac{5}{14},$$
$$P(S_2) = P(S_3) = \frac{2}{14}.$$
▲

برای تکمیل بحث در این‌جا یادآوری می‌کنیم که گاهی اوقات نمودارهای حالت با ***نمودارهای داربستی*** جایگزین می‌شوند. در واقع نمودار داربستی یک نمودار حالت است که با محور زمان افزایش می‌یابد به قسمی که تغییرات حالت را می‌توان به صورت تابعی از زمان دید. شکل (۴.۳- الف و ب) را مقایسه کنید. در این صورت هر زنجیر مارکوف با مسیر خاصی در نمودار داربستی متناظر می‌باشد.

چند ویژگی زنجیر مارکوف را در این فصل ذکر خواهیم کرد. در این‌جا دو ویژگی که کاربردهایی در نظریهٔ اطلاع پیدا می‌کنند بدون نتیجه یادآوری شده‌اند:

(الف) بخشی از زنجیر مارکوف نیز یک زنجیر مارکوف است.
(ب) زنجیر مارکوفی که از جهت عکس عبور کند نیز یک زنجیر مارکوف است.



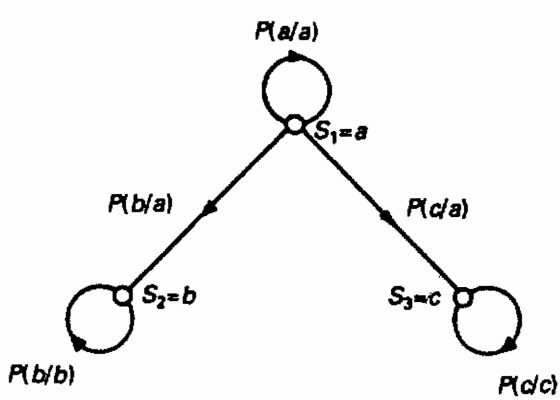
الف ب

شکل ۴.۳- (الف) نمودار حالت؛ (ب) نمودار داربستی

در این‌جا محدودیتهایی بر زنجیرهای مارکوف مورد بررسی تحمیل خواهد شد. در مرحلهٔ اوّل نیاز داریم که در هر انتقال ماتریس احتمالهای انتقال یکسان باشد. در این صورت احتمالهای انتقال را *مانا* و زنجیر مارکوف را **همگن** می‌نامیم. به علاوه لازم است زنجیر مارکوف خودش نیز *مانا* باشد؛ یعنی این‌که احتمالهای حالتهایی که زنجیر مارکوف می‌تواند به آن برسد تغییر نکند. نهایتاً، علاوه بر این می‌خواهیم خود را به زنجیرهای ***ارگودیک*** محدود کنیم؛ این بدین معنی است که مهم نیست که زنجیر در چه حالتی قرار دارد، سرانجام از هر حالتی می‌تواند به هر حالت دیگری برسد.

۲.۳ اطلاع یک منبع گسسته باحافظه

در حالت منابع گسسته با حافظه مقدار معیّنی وابستگی بین نمادهای متوالی وجود دارد. وابستگی که در این‌جا اشاره شد می‌تواند بر دنباله‌های طویل دلخواهی از نمادها گسترش یابد. با وجود این، اغلب می‌توان فرض کرد که این وابستگی بر تعداد محدودی از نمادها گسترش می‌یابد، برای این که کاربرد زنجیر مارکوف تعریف شده در بخش (۱.۳) را به عنوان الگویی برای منبع اطلاع با حافظه ممکن سازد. مرتبهٔ زنجیر مارکوف را به طور تجربی می‌توان تعیین کرد. روشهای برآوردی برای این کار وجود دارد، ولی چون خارج از حوصلهٔ این کتاب است در این‌جا اشارهٔ بیشتری به آن نخواهد شد. از این پس فرض می‌کنیم که منبع اطلاع ارگودیک است. این امر موجب می‌شود که در زمان معیّنی در مورد احتمال نماد معیّنی اظهار نظر کنیم. در این صورت این احتمال برابر است با احتمال این‌که این نماد در دنباله طویلی (بی‌نهایت) از نمادهای متوالی ظاهر می‌شود. برای مثال یک



شکل ۵.۳- زنجیر مارکوف غیرارگودیک

منبع اطلاع گسسته با حافظه به عنوان الگویی برای نوشته‌ها عمل می‌کند و همین‌طور برای هر دنبالهٔ تصادفی که از آن نتیجه شود. مثال دیگر سیگنالی است که از لحاظ نمونه‌گیری تعیین مقدار شده است، که در آن فراوانی نمونه‌گیری به طور صحیح انتخاب نشده است و بنابراین سبب استقلال بین نمونه‌ها (تعیین مقدار شده) می‌گردد.

I مقدار اطلاع برای زنجیر مارکوف مرتبه-اوّل

برای یک زنجیر مارکوف مرتبه-اوّل همچنان که در بخش قبل دیدیم تعداد نمادهای u_i، $i = ۱,۲,\ldots,m$، برابر تعداد حالتهای S_i می‌باشد. اکنون انتقالهایی از یک نماد دلخواه $u_{۱i}$ در لحظه $t_۱$ به نماد $u_{۲j}$ در لحظهٔ $t_۲$، $i,j = ۱,۲,\ldots,m$، را در نظر خواهیم گرفت. احتمال شرطی، احتمال انتقال از $u_{۱i}$ به $u_{۲j}$، را می‌توان با $p(u_{۲j}|u_{۱i})$ نشان داد. اکنون مقدار اطلاع متعلّق به یک انتقال دلخواه به صورت زیر داده می‌شود (با تعریف (۳.۱) مقایسه کنید)

$$H(U_۲|U_۱) = -\sum_{i=۱}^{m}\sum_{j=۱}^{m} P(u_{۱i},u_{۲j})\log P(u_{۲j}|u_{۱i}). \qquad (۴.۳)$$

برای مقدار اطلاع توأم دو نماد داریم (با تعریف (۲.۱) مقایسه کنید)

$$H(U_۱,U_۲) = -\sum_{i=۱}^{m}\sum_{j=۱}^{m} P(u_{۱i},u_{۲j})\log P(u_{۱i},u_{۲j}). \qquad (۵.۳)$$

و

$$H(U_۱,U_۲) = H(U_۱) + H(U_۲|U_۱). \qquad (۶.۳)$$

بنابراین مقدار اطلاع در یک پیام با طول دو با مجموع مقدار اطلاع نماد اوّل و مقدار اطلاع شرطی نماد دوم به شرط نماد اوّل برابر است. همان‌طور که در قضیهٔ (۲.۱) به دست آوردیم، داریم

$$H(U_۲|U_۱) \le H(U_۲), \qquad (۷.۳)$$

و بنابراین از معادلهٔ (۶.۳) نتیجه می‌شود که

$$H(U_۱,U_۲) \le H(U_۱) + H(U_۲). \qquad (۸.۳)$$

اگر نمادهای متوالی به طور آماری مستقل باشند، بدین معنی که اگر منبع بی‌حافظه باشد برابری برقرار است. چون منبع مانا و ارگودیک است، $H(U_۱) = H(U_۲) = H(U)$، بنابراین می‌توان نوشت

$$H(U_۱,U_۲) \le ۲H(U). \qquad (۹.۳)$$

پس مقدار اطلاع در یک پیام شامل دو نماد برای یک منبع با حافظه کوچکتر از یک منبع بی‌حافظه است.

II مقدار اطلاع برای زنجیرهای مارکوف مرتبه-بالاتر

چون یک منبع اطلاع گسسته دلخواه می‌تواند یک زنجیر مارکوف از مرتبهٔ $k>1$ تولید کند، پسندیده آن است که مطالب گفته شده در بالا را به منابعی با حافظه زیاد دلخواه گسترش دهیم. مقدار اطلاع شرطی $F_N(U)$ نماد u_N را در حالتی که $N-1$ نماد قبلی معلوم‌اند بررسی خواهیم کرد، یعنی

$$F_N(U)=H(U_N|U_{N-1},\ldots,U_2,U_1). \tag{۱۰.۳}$$

این مقدار اطلاع شرطی چند ویژگی دارد. اوّلین ویژگی عبارت است از:

$$H(U_N|U_{N-1},\ldots,U_2,U_1)\le H(U_N|U_{N-1},\ldots,U_2). \tag{۱۱.۳}$$

در واقع آنچه که این ویژگی بیان می‌کند این است که اطلاعی که از نماد اوّل به دست می‌آید نمی‌تواند موجب افزایش عدم‌حتمیت دربارهٔ نماد Nام گردد، ولی آن را کاهش می‌دهد یا بدون تغییر باقی می‌گذارد.

قضیهٔ ۱.۳

مقدار اطلاع شرطی $F_N(U)=H(U_N|U_{N-1},\ldots,U_1)$ نماد Nام در حالی که $N-1$ نماد قبلی معلوم‌اند تابع نزولی یکنواختی از N است، یعنی:

$$\begin{aligned}H(U_N|U_{N-1},\ldots,U_1)\le H(U_{N-1}|U_{N-2},\ldots,U_1)\le\cdots\\ \cdots\le H(U_2|U_1)\le H(U_1).\end{aligned} \tag{۱۲.۳}$$

برهان

چون منبع ماناست، مقادیر اطلاع شرطی مستقل از مکان نماد Nام در زنجیر می‌باشد. از این‌رو برای مثال داریم

$$H(U_{N-1}|U_{N-2},\ldots,U_1)=H(U_N|U_{N-1},\ldots,U_2),$$

و بنابر ویژگی اوّل (معادله (۱۱.۳) را ببینید) مستقیماً نتیجه می‌شود که:

$$H(U_N|U_{N-1},\ldots,U_1)\le H(U_{N-1}|U_{N-2},\ldots,U_1),$$

بنابراین

$$F_N(U)\le F_{N-1}(U)\le\cdots\le F_2(U)\le F_1(U). \quad ■$$

این قضیه بیان می کند که با افزایش مقدار N، مقدار اطلاع شرطی کوچکتر می‌شود یا حداّکثر به همان اندازه باقی می‌ماند. چون هر مقدار اطلاع همواره بزرگتر یا برابر صفر است، نتیجه می‌شود که $F_N(U)$ به یک مقدار حدّی میل می کند، که آن را به صورت زیر نمایش می‌دهیم

$$H_\infty(U) = \lim_{N\to\infty} F_N(U) = \lim_{N\to\infty} H(U_N|U_{N-1},\ldots,U_1) \text{ بیت / نماد.} \qquad (۱۳.۳)$$

بدیهی است که اگر الفبای منبع U شامل m نماد باشد، داریم

$$0 \le H_\infty(U) \le \log m. \qquad (۱۴.۳)$$

اکنون مقدار $H_\infty(u)$ را به عنوان مقدار اطلاع یک منبع اطلاع گسسته با حافظه تعریف می کنیم؛ بنابراین حافظه ممکن است با طول نامحدود باشد. اگر منبع زنجیر مارکوفی از مرتبه k تولید کند در این صورت بدین معناست که

$$P(u_N|u_{N-1},\ldots,u_1) = P(u_N|u_{N-k},\ldots,u_{N-1}). \qquad (۱۵.۳)$$

در نتیجه برای مقدار اطلاع شرطی داریم

$$\begin{aligned} H(U_N|U_{N-1},\ldots,U_1) &= H(U_N|U_{N-k},\ldots,U_{N-1}) \\ &= H(U_{k+1}|U_k,\ldots,U_1). \end{aligned} \qquad (۱۶.۳)$$

از این رو با افزایش N هیچ افزایشی در طول حافظه به وجود نمی‌آید زیرا محدود به k باقی می‌ماند. این بدین معناست که از $N = k+1$ به بعد کمیّت $F_N(U)$ برابر با $F_{k+1}(U)$ باقی می‌ماند و از این رو بیشتر از این کاهش نمی‌یابد. در این صورت نتیجه می‌شود که مقدار حدّ $H_\infty(U)$ برای زنجیر مارکوف از مرتبهٔ k برابر است با

$$H_\infty(U) = F_{k+1}(U) = H(U_{k+1}|U_k,\ldots,U_2,U_1). \qquad (۱۷.۳)$$

اگر منبع بی‌حافظه باشد در این صورت $k=0$، بنابراین $H_\infty(U)$ با $H(U)$ برابر است. با افزایش مرتبهٔ k، $H_\infty(U)$ دائماً کوچکتر خواهد شد.

برای مثال علاوه بر نمادهای تکی همچنین اغلب پیامهایی که از N نماد ساخته شده‌اند بررسی خواهد شد. می‌توان مقدار اطلاع هر نماد را بر اساس مقدار اطلاع هر پیام $H(V)$ به دست آورد. کمیت $H(V)$ به صورت زیر تعریف می‌شود:

$$H(V) = H(U_1,U_2,\ldots,U_N) \text{ بیت / پیام.} \qquad (۱۸.۳)$$

اکنون مقدار اطلاع برای هر نماد به صورت زیر تعریف می شود

$$H_N(U)=\frac{1}{N}H(V)=\frac{1}{N}H(U_1,U_2,\ldots,U_N)\ \text{نماد / بیت}. \tag{۱۹.۳}$$

اگر نمادهای u_i به طور آماری مستقل باشند، در این صورت داریم

$$H_N(U)=\frac{1}{N}\sum_{i=1}^{N}H(U_i)=\frac{1}{N}N\,H(U)=H(U).$$

اگر نمادها وابسته باشند، در این صورت داریم

$$H_N(U)=\frac{1}{N}[H(U_1)+H(U_2|U_1)+\cdots+H(U_N|U_{N-1},\ldots,U_2,U_1)]$$

$$=\frac{1}{N}\sum_{j=1}^{N}F_j(U). \tag{۲۰.۳}$$

همان‌طور که برای $F_N(U)$ داشتیم، $H_N(U)$ نیز کاهشی یکنواخت است و با افزایش N به مقدار حدّ $H_\infty(U)$ می‌رسد.

قضیهٔ ۲.۳

اگر $H(V)$ مقدار اطلاع برای یک پیام به طول N باشد آن‌گاه مقدار اطلاع برای هر نماد تعریف شده با $H_N(U)=\frac{H(V)}{N}$، کاهشی یکنواخت است. به علاوه داریم

$$\lim_{N\to\infty} H_N(U)=H_\infty(u). \tag{۲۱.۳}$$

برهان

با بهره‌وری از معادلات (۱۲.۳) و (۲۰.۳) برای $H(V)$ نتیجه می‌شود که

$$H(V)=N\,H_N(U)=H(U_1)+H(U_2|U_1)+\cdots+H(U_N|U_{N-1},\ldots,U_2,U_1)$$

$$\geq N\,H(U_N|U_{N-1},\ldots,U_2,U_1),$$

یا با فرمول (۱۰.۳)داریم

$$H_N(U)\geq F_N(U). \tag{۲۲.۳}$$

اکنون می‌توان نوشت

$$H(V)=H(U_1,\ldots,U_N)=H(U_1,\ldots,U_{N-1})+H(U_N|U_{N-1},\ldots,U_1)$$

یا

$$N\,H_N(U)=(N-1)H_{N-1}(U)+F_N(U)$$

$$\leq(N-1)H_{N-1}(U)+H_N(U).$$

بنابراین داریم

$$(N-1)H_N(U) \le (N-1)H_{N-1}(U),$$

یا

$$H_N(U) \le H_{N-1}(U), \qquad (۲۳.۳)$$

این ثابت می‌کند که $H_N(U)$ کاهشی یکنواخت است. چون $H_N(U) \ge 0$، باید به یک حد میل کند. این حد نیز برابر $H_\infty(U)$ است. همان‌طور که قبلاً نشان داده شد

$$H_N(U) = \frac{1}{N}\sum_{j=1}^{N} F_j(U).$$

چون وقتی $j \to \infty$، $F_j(U)$ به $H_\infty(U)$ میل می‌کند نتیجه می‌شود که

$$\lim_{N\to\infty} H_N(U) = \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} F_j(U) = \frac{1}{N}[N\,H_\infty(U)] = H_\infty(U),$$

که با معادلهٔ (۲۱.۳) موافق می‌باشد. ■

همچنان که از معادلهٔ (۱۳.۳) و قضیهٔ (۲.۳) نتیجه می‌شود، هر دو تابع $F_N(U)$ و $H_N(U)$ به یک حد میل می‌کنند. همان‌طور که نشان داده شده است $H_N(U) \ge F_N(U)$، بنابراین $H_N(U)$ یک تقریب نادرست از مقدار اطلاع واقعی $H_\infty(U)$ است. با وجود این، یک مزیّت $H_N(U)$ سادگی آن است.

مثال ۲.۳

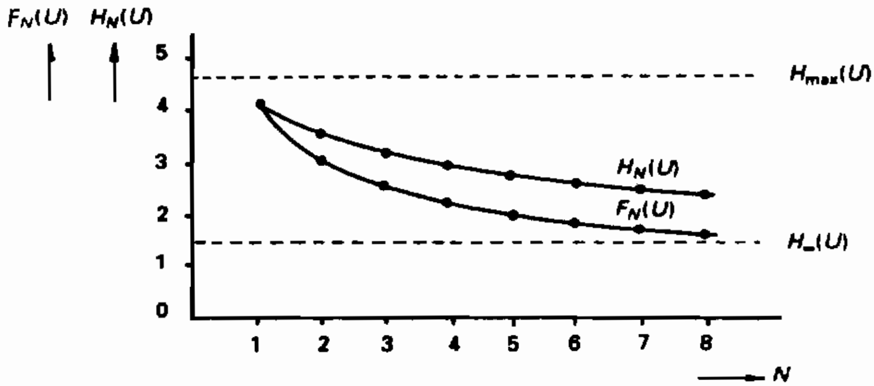
فرض کنید مقادیر مختلف $F_j(U)$ برای ۲۶ نماد مختلف معلومند (برای مثال مربوط به زبان)

$F_1 = ۴٫۱۵$	آن‌گاه	$H_1 = F_1$	$= ۴٫۱۵$
$F_2 = ۲٫۹۹$		$H_2 = (1/2)(F_1 + F_2)$	$= ۳٫۷۵$
$F_3 = ۲٫۵۶$		$H_3 = (1/3)(F_1 + F_2 + F_3)$	$= ۳٫۲۳$
$F_4 = ۲٫۲۰$		$H_4 = (1/4)(F_1 + \cdots)$	$= ۲٫۹۸$
$F_5 = ۱٫۹۵$		$H_5 = ---$	$= ۲٫۷۷$
$F_6 = ۱٫۷۲$		$H_6 = ---$	$= ۲٫۶۰$
$F_7 = ۱٫۶۳$		$H_7 = ---$	$= ۲٫۴۶$
$F_8 = ۱٫۶۰$		$H_8 = ---$	$= ۲٫۳۵$
-----		---	----

از شکل (۶.۳) نتیجه می‌شود که حدّ $F_N(U)$ و $H_N(U)$ تقریباً برابر است با

$H_\infty(U)=$ ۱٫۵۰ بیت / نماد ،

در حالی که $\max_u H(U)=\log$ ۲۶ $=$ ۴٫۷۰ بیت / نماد. ▲



شکل ۳-۶. $F_N(U)$ و $H_N(U)$ به عنوان تابعی از N

۳.۳ جنبه‌های کدگذاری

مانند حالت منبع بی‌حافظه می‌توان بزرگی تعداد محتملترین پیامهایی را که یک منبع با حافظه می‌تواند تولید کند تعیین کرد. باید انتظار داشت که به واسطهٔ حافظه این عدد کوچکتر از عدد یک منبع بی‌حافظه باشد. این حقیقتاً ثابت می‌کند که این چنین است. می‌توان ثابت کرد که اگر به جای N تعداد ℓ نماد را با یکدیگر انتخاب کنیم، در این صورت با افزایش ℓ، $\frac{-\log p(v)}{\ell}$ به $H_\infty(U)$ میل می‌کند.

قضیهٔ ۳.۳

برای هر $\varepsilon>0$ و $\delta>0$ معلوم می‌توان ℓ_0 را به قسمی یافت که دنباله‌هایی از هر طول $\ell\geq\ell_0$ به دو دسته تقسیم شوند:

(الف) مجموعهٔ S' که احتمال کلّ آن کمتر از ε است.

(ب) مجموعهٔ باقی‌ماندهٔ S، همهٔ اعضایی که احتمال آنها در نابرابری زیر صدق می‌کنند

$$\left|\frac{-\log p(v)}{\ell}-H_\infty(U)\right|<\delta. \qquad (۲۴.۳)$$

این گروه، مجموعهٔ ***محتملترین پیامهاست***. ■

این قضیه نظیر قضیهٔ شانون-مک‌میلان (قضیهٔ (۳.۲) را ببینید) است که مربوط به منبع بی‌حافظه است. بدیهی است که در این حالت معادلهٔ (۲۴.۳) با معادلهٔ (۲۴.۲) یکی می‌شود چون در این صورت داریم $H_\infty(U)=H(U)$.

گروه S محتملترین پیامها با احتمال $P(S)>1-\varepsilon$ وجود دارد که در آن هر پیام دارای احتمال زیر است

$$P(v)\approx 2^{-l.H_\infty(U)}. \tag{۲۵.۳}$$

تعداد محتملترین پیامها تقریباً برابر است:

$$M_\infty=\frac{1}{p(v)}\approx 2^{l.H_\infty(U)}. \tag{۲۶.۳}$$

چون $H_\infty(U)\le H(U)$، M_∞ تعداد محتملترین پیامهای یک منبع با حافظهٔ کوچکتر یا برابر با تعداد محتملترین پیامهای یک منبع بی‌حافظه است.

در بخش (۱.۲) **حشو** به صورت زیر تعریف شد

$$red=1-\frac{H(U)}{\max\limits_u H(U)}=1-\frac{H(U)}{\log n}. \tag{۲۷.۳}$$

از این رو این اثر کیفیت منبع با حافظه را اندازه‌گیری می‌کند. در این فصل، نتیجه می‌شود که بستگی بین نمادها نیز موجب مقدار اطلاع کمتری می‌شود. می‌توان این اتلاف اطلاع را با ***حشو وابسته*** نشان داد

$$red_\infty=1-\frac{H_\infty(U)}{H(U)}, \tag{۲۸.۳}$$

که در آن $H_\infty(U)$ مقدار اطلاع منبع با حافظه و $H(U)$ مقدار اطلاع منبع بی‌حافظه است که نمادهای آن با نمادهای منبع با حافظه احتمالهای یکسانی دارند. در خاتمه، تعریف ***حشو کل*** را خواهیم داد

$$red_{\text{کل}}=1-\frac{H_\infty(U)}{\max\limits_u H(U)}=1-\frac{H_\infty(U)}{\log n}. \tag{۲۹.۳}$$

مثال ۳.۳

در مثال (۲.۳) نماد / بیت $\max H(U)=۴{,}۷۰$، نماد / بیت $H(U)=۴{,}۱۵$ و نماد / بیت $H_\infty(U)=۱{,}۵۰$، برای اندازه‌های حشو مختلف مقادیر زیر را به دست می‌آوریم

$$red=1-\frac{۴{,}۱۵}{۴{,}۷۰}=۰{,}۱۲,$$

$$red_\infty = ۱ - \frac{۱٫۵۰}{۴٫۱۵} = ۰٫۶۴,$$

و

$$red_{کل} = ۱ - \frac{۱٫۵۰}{۴٫۷۰} = ۰٫۶۸. \qquad \blacktriangle$$

از این رو نسبت به منابع باحافظه سه اندازهٔ مختلف حشو وجود دارد و هر یک بعضی ویژگی خاص وابستگی واقعی بین نمادها را بیان می‌کند. برای منبع مارکوف مرتبهٔ صفر داریم $red = red_{کل}$ و $red_\infty = ۰$.

می‌توان یک قضیهٔ کدگذاری منبع برای یک منبع اطلاع باحافظه مشابه قضیهٔ اوّل کدگذاری شانون به دست آمده در بخش (۴.۲) (قضیهٔ ۵.۲) را با قرار دادن $H_\infty(U)$ به جای $H(U)$ به دست آورد. چون اثبات آن مانند قضیهٔ (۵.۲) است در این جا نخواهیم آورد.

قضیهٔ ۴.۳

برای یک منبع اطلاع گسسته باحافظه و مقدار اطلاع $H_\infty(U)$، که در آن پیامهای به طول ℓ در کدواژه‌هایی به طول L از یک الفبای کد با حجم r رمزی شده‌اند، پیامی وجسود دارد که برای آن هیچ کدواژه‌ای وجود ندارد که بتوان P_e احتمال آن را به دلخواه کوچک ($P_e < \varepsilon$) کرد، اگر L در نابرابری زیر صدق کند

$$L \log r \geq \ell H_\infty(U), \qquad (۳۰.۳)$$

و ℓ به قدر کافی بزرگ باشد.

همان‌طور که از مثال (۳.۳) نتیجه می‌شود حشو وابسته می‌تواند بین بقیه به قدر کافی بزرگ باشد که در این صورت مطلوب آن است که از طریق کدگذاری حذف شود. برای یک منبع بی‌حافظه نشان داده شده است که چگونه روشهای کدگذاری برای به دست آوردن یک کد مناسب که متوسط طول کدواژه را می‌نیمم کند توسعه داد به طوری که هم‌زمان حشو در کدنمادها را می‌نیمم کند. یک راه حذف کردن حشو وابسته به منبع با حافظه این است که این روشهای کدگذاری را به جای نمادهای تکی بر پیامهایی با طول l به کار بریم. از این رو این رهیافت برمبنای گسترش الفبا می‌باشد. می‌توان طول l را با مرتبهٔ زنجیر مارکوف انتخاب کرد.

مثال ۴.۳

یک منبع اطلاع را که یک زنجیر مارکوف مرتبهٔ اوّل تولید می‌کند در نظر بگیرید. الفبای منبع عبارت است از $U = \{A,B,C\}$؛ احتمالهای انتقال زیر داده شده‌اند:

$$p(A|A)=\frac{۱}{۲},\qquad p(B|A)=\frac{۱}{۲},\qquad p(C|A)=۰,$$
$$p(A|B)=\frac{۱}{۴},\qquad p(B|B)=۰,\qquad p(C|B)=\frac{۳}{۴},$$
$$p(A|C)=\frac{۱}{۳},\qquad p(B|C)=\frac{۱}{۳},\qquad p(C|C)=\frac{۱}{۳},$$

احتمالهای حاشیه‌ای از معادلات زیر نتیجه می‌شود.

$$\begin{cases} p(A)=\frac{۱}{۲}p(A)+\frac{۱}{۴}p(B)+\frac{۱}{۳}p(C),\\ p(B)=\frac{۱}{۲}p(A)+\qquad\qquad\frac{۱}{۳}p(C),\\ p(C)=\qquad\qquad\frac{۳}{۴}p(B)+\frac{۱}{۳}p(C),\\ p(A)+p(B)+p(C)=۱. \end{cases}$$

نتایج عبارتند از $p(A)=\frac{۱۰}{۲۷}$، $p(B)=\frac{۸}{۲۷}$ و $p(C)=\frac{۹}{۲۷}$.

فرض کنید درست دو کدنماد ترکیب شده باشند، براساس احتمالهای انتقال احتمالهای توأم نظیر را می‌توان یافت. در جدول بعد این احتمالهای توأم با کدواژه‌های به دست آمده از کاربرد روش فانو برای کدگذاری برای حالت $r=۲$ داده شده‌اند.

متوسط طول کدواژه برابر است با $L=\frac{۷۵}{۲۷}\approx ۲٫۷۸$ یا ۱٫۳۹ بر نماد. با محاسبهٔ مقدار اطلاع توأم به دست می‌آوریم: $H(U_۱,U_۲)=۲٫۷۲$. بنابراین کارایی برابر است با $\eta=\frac{H(U_۱,U_۲)}{L}=\frac{۲٫۷۲}{۲٫۷۸}\approx ۰٫۹۸$.

	احتمال	کدواژه
BC	۶/۲۷	۰۰
AA	۵/۲۷	۰۱
AB	۵/۲۷	۱۰۰
CA	۳/۲۷	۱۰۱
CB	۳/۲۷	۱۱۰
CC	۳/۲۷	۱۱۱۰
BA	۲/۲۷	۱۱۱۱
AC	۰	–
BB	۰	–

اگر روش فانو را برای نمادهای مجزّای منبع به کار بریم متوسط طول کدواژه برابر $L = \frac{۴۴}{۲۷} \approx ۱٫۶۳$ خواهد شد، که بزرگتر از ۱٫۳۹ بر نماد داده شده در بالا می‌باشد. چنان که می‌توان نشان داد کارایی به $\eta = ۰٫۹۷$ کاهش می‌یابد. ▲

همچنین می‌توان منبع اطلاع را به روش دیگری تنظیم کرده آن را به یک منبع بدون اطلاع برگشت داد. یک مثال از این نوع به نام طول گردش معروف است که در میان دیگر روشها برای کدگذاری مدارک به کار برده می‌شود. با این روش مدرک را خط به خط به طور اجمالی بررسی نموده به نقاط خیالی سفید و سیاه رقم‌گذاری می‌شود. در این روش یک منبع اطلاع دودویی باحافظه به دست می‌آید که دو نماد ۰ و ۱ متناظر با سفید و سیاه تولید می‌کند. به طور تقریبی می‌توان فرض کرد که بستگی بین نمادها را می‌توان با یک زنجیر مارکوف مرتبهٔ اوّل مشخّص کرد که در آن $p(۰)$ احتمال نقطهٔ خیالی سفید به طور قابل ملاحظه‌ای بزرگتر از $p(۱)$ احتمال نقطهٔ خیالی سیاه است. اکنون به ترتیب دنباله‌های صفرها و یکهای متوالی را بررسی می‌کنیم و طول این دنباله‌ها یا گردشها را تعیین می‌کنیم. هر دنباله‌ای از k صفر را که به یک ختم می‌شود، می‌توان به صورت دنباله‌ای از $k-۱$ انتقال از سفید به سفید که توسط یک انتقال از سفید به سیاه دنبال می‌شود در نظر گرفت. در این صورت احتمال چنین دنباله‌ای عبارت است از

$$P_k(۰) = P(۰|۰)^{k-۱} \cdot P(۱|۰). \qquad (۳۱.۳)$$

به طور مشابه، احتمال دنباله‌ای از k یک (نقاط سیاه) برابر است با

$$P_k(۱) = P(۱|۱)^{k-۱} \cdot p(۰|۱). \qquad (۳۲.۳)$$

این را می‌توان به کاربرد و به ترتیب متوسط طول دنباله‌ای از سفید یا سیاه را تعیین کرد. در این صورت به دست می‌آوریم

$$\overline{k(۰)} = \sum_{k=۱}^{\infty} k\, P(۰|۰)^{k-۱} \cdot P(۱|۰) = \frac{۱}{P(۱|۰)}, \qquad (۳۳.۳)$$

$$\overline{k(۱)} = \sum_{k=۱}^{\infty} k\, P(۱|۱)^{k-۱} \cdot p(۰|۱) = \frac{۱}{p(۰|۱)}. \qquad (۳۴.۳)$$

اکنون منبع اطلاع اصلی را به صورت یک منبع که کلّ اعداد را تولید می‌کند، یعنی طولهای گردش با احتمالهای $P_k(۰)$ و $P_k(۱)$ برای $k = ۱,۲,\ldots,\infty$ در نظر می‌گیریم. به ویژه، یک طول ماکسیمم یعنی K را جایز دانسته و دنباله‌های طولانی‌تر را مضربی از K در نظر

می‌گیریم. در پایان، می‌توان برای مثال کد هافمن را برای کدگشایی این طولهای گردشی به کار برد.

۴.۳ تمرینها

۱.۳ یک منبع اطلاع دارای الفبای $\{u_۱, u_۲, u_۳\}$ است و یک زنجیر مارکوف مرتبه اوّل را تولید می‌کند. احتمالهای انتقال به صورت زیر داده شده‌اند:

$$P(u_۱|u_۱)=\frac{۱}{۲},\quad P(u_۲|u_۱)=\frac{۱}{۲},\quad P(u_۳|u_۱)=۰,$$
$$P(u_۱|u_۲)=\frac{۱}{۳},\quad P(u_۲|u_۲)=۰,\quad P(u_۳|u_۲)=\frac{۲}{۳},$$
$$P(u_۱|u_۳)=\frac{۱}{۳},\quad P(u_۲|u_۳)=\frac{۲}{۳},\quad P(u_۳|u_۳)=۰.$$

(الف) نمودار حالتهای متعلّق به زنجیر مارکوف را رسم کنید. آیا زنجیر مارکوف ارگودیک است.

(ب) احتمالهای نمادهای $u_۱$، $u_۲$ و $u_۳$ را بیابید.

۲.۳ یک منبع اطلاع با الفبای $\{۰,۱\}$ یک زنجیر مارکوف مرتبهٔ دوم تولید می‌کند، که با احتمالهای انتقال زیر توصیف شده‌اند:

$$P(۰|۰۰)=۰٫۸,\quad P(۰|۱۱)=۰٫۲,$$
$$P(۱|۰۰)=۰٫۲,\quad P(۱|۱۱)=۰٫۸,$$
$$P(۰|۰۱)=۰٫۵,\quad P(۰|۱۰)=۰٫۵,$$
$$P(۱|۰۱)=۰٫۵,\quad P(۱|۱۰)=۰٫۵.$$

(الف) نمودار حالت متعلّق به این زنجیر را رسم کنید.

(ب) احتمالهای حالتهای S_i را بیابید.

۳.۳ یک منبع اطلاع با الفبای منبع $\{۰,۱\}$ یک زنجیر مارکوف مرتبهٔ دوم تولید می‌کند که با احتمالهای انتقال زیر مشخّص شده است:

$$P(۰|۰۰)=\frac{۱}{۴},\quad P(۰|۰۱)=\frac{۱}{۴},$$
$$P(۰|۱۰)=\frac{۳}{۴},\quad P(۰|۱۱)=\frac{۳}{۴}.$$

(الف) نشان دهید که زنجیر مارکوف با احتمالهای انتقال داده شده کاملاً توصیف شده است.
(ب) نمودار حالت را رسم کنید.
(پ) احتمالهای هر یک از حالتها را محاسبه کنید.
(ت) احتمالهای نمادهای منبع خروجی را بیابید.
(ث) نشان دهید که چگونه می‌توان با تغییر یکی از احتمالهای انتقال داده شده زنجیر غیرارگودیک ساخت.

۴.۳ یک منبع اطلاع دارای الفبای $\{u_۱,u_۲,u_۳\}$ است و نمادهایی که یک زنجیر مارکوف مرتبهٔ اوّل می‌سازد تولید می‌کند. احتمالهای انتقال به صورت زیر می‌باشند:

$$\begin{aligned} &P(u_۱|u_۱)=۰, \qquad &&P(u_۲|u_۱)=\frac{۱}{۵}, \qquad &&P(u_۳|u_۱)=\frac{۴}{۵},\\ &P(u_۱|u_۲)=\frac{۱}{۲}, &&P(u_۲|u_۲)=\frac{۱}{۱۰}, &&P(u_۳|u_۲)=\frac{۲}{۵},\\ &P(u_۱|u_۳)=\frac{۱}{۲}, &&P(u_۲|u_۳)=۰, &&P(u_۳|u_۳)=\frac{۱}{۲}. \end{aligned}$$

(الف) نمودار حالت را رسم کنید.
(ب) احتمالهای نمادهای $u_۱$، $u_۲$ و $u_۳$ را محاسبه کنید.
(پ) مقدار اطلاع یک منبع باحافظه را که نمادهایش دارای احتمال مساوی به عنوان نمادهایی از منبع در نظر گرفته شده‌اند حساب کنید.
(ت) مقدار اطلاع همراه با یک انتقال دلخواه از منبع باحافظه را بیابید.
(ث) مقدار اطلاع توأم دو نماد را محاسبه کنید.
(ج) حشو، حشو وابسته و حشو کل را محاسبه کنید.
(چ) رابطه‌ای برای حشو کل برحسب تعداد محتملترین پیامها به دست آورید.

۵.۳ یک منبع اطلاع دارای الفبای $\{u_۱,u_۲,u_۳\}$ است و یک زنجیر مارکوف مرتبهٔ اوّل مانا تولید می‌کند. احتمالهای انتقال از یک نماد u_i به نماد u_j با $i \neq j$ همگی برابر $\frac{p}{۲}$ هستند.

(الف) نمودار حالت این زنجیر مارکوف را رسم کنید.
(ب) احتمالهای نمادهای $u_۱$، $u_۲$ و $u_۳$ را بیابید.
(پ) مقدار اطلاع نسبت به یک انتقال دلخواه را محاسبه کنید.

(ت) تعیین کنید برای چه مقدار p این مقدار اطلاع ماکسیمم می‌شود.

(ث) آیا چه مفهومی به مقادیر $H(U)$ به دست‌آمده برای $p=۰$ و $p=۱$، و به ماکسیمم مقدار $H(U)$ نسبت می‌دهید؟

۶.۳ یک منبع اطلاع دو نماد ۰ و ۱ را تولید می‌کند. دنباله نمادهای تولید شده یک زنجیر مارکوف مرتبهٔ دوم با احتمالهای انتقال زیر می‌سازد:

$$P(۰|۰۰) = ۰{,}۸,$$
$$P(۰|۰۱) = ۰{,}۵,$$
$$P(۰|۱۰) = ۰{,}۵,$$
$$P(۰|۱۱) = ۰{,}۲.$$

(الف) مقدار اطلاع یک سه‌تایی ایجاد شده از این منبع اطلاع چه قدر است؟ با استفاده از

$$p(۰۰) = p(۱۱) = \frac{۵}{۱۴},$$
$$p(۰۱) = p(۱۰) = \frac{۱}{۷}.$$

با بهره‌وری از این نتیجه، مقدار اطلاع برای هر نماد را که با $H_۳(U)$ نشان داده می‌شود پیدا کنید.

(ب) مقدار اطلاع یک دوتایی چه‌قدر است؟ بنابراین $H_۲(U)$ را پیدا کنید.

(پ) $H_۱(U)$ چه قدر است؟

(ت) مقدار اطلاع شرطی در پیشامدی که $N-۱$ نماد قبلی داده شده‌اند با $F_N(U)$ نشان داده می‌شود. $F_۱(U)$، $F_۲(U)$ و $F_۳(U)$ را تعیین کنید.

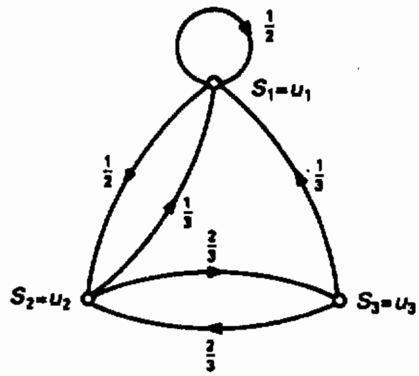
(ث) شرح دهید چرا $F_۳(U) < F_۲(U) < F_۱(U)$ ؟

(ج) دربارهٔ مقادیر $F_۲(U)$ و $H_۲(U)$ چه می‌توانید بگویید؟

(چ) $F_N(U)$ و $H_N(U)$ را به عنوان تابعی از N تعداد نمادها رسم کنید.

۵.۳ جوابها

۱.۳ (الف) زنجیر مارکوف از مرتبهٔ $k=۱$ است، به قسمی که تعداد حالتها برابر $m^k = ۳$ است. در این صورت نمودار حالت به صورتی است که در شکل (۷.۳) نشان داده شده است.



شکل ۷.۳- نمودار حالت تمرین (۱.۳)

زنجیر ارگودیک است چون از هر حالت می‌توان به حالت دیگر رسید.

(ب) محاسبهٔ احتمال نمادهای $u_۱$، $u_۲$ و $u_۳$ متناظر است با محاسبهٔ احتمال سه حالت $S_۱$، $S_۲$ و $S_۳$ که از معادلات زیر نتیجه می‌شود

$$P(u_۱) = P(u_۱).P(u_۱|u_۱) + P(u_۲).P(u_۱|u_۲) + P(u_۳).P(u_۱|u_۳),$$

$$P(u_۲) = P(u_۱).P(u_۲|u_۱) + P(u_۲).P(u_۲|u_۲) + P(u_۳).P(u_۲|u_۳),$$

$$P(u_۳) = P(u_۱).P(u_۳|u_۱) + P(u_۲).P(u_۳|u_۲) + P(u_۳).P(u_۳|u_۳).$$

علاوه براین شرط زیر نیز باید برقرار باشد

$$P(u_۱) + P(u_۲) + P(u_۳) = ۱.$$

با قرار دادن احتمالهای داده شده در این معادلات نتیجه می‌شود:

$$\begin{aligned} P(u_۱) &= \frac{۱}{۲}P(u_۱) + \frac{۱}{۳}P(u_۲) + \frac{۱}{۳}P(u_۳), \\ P(u_۲) &= \frac{۱}{۲}P(u_۱) \qquad\qquad + \frac{۲}{۳}P(u_۳), \\ P(u_۳) &= \qquad\qquad \frac{۲}{۳}P(u_۲), \\ ۱ &= P(u_۱) + P(u_۲) + P(u_۳). \end{aligned}$$

با حلّ این چهار معادلهٔ سه مجهولی به دست می‌آوریم

$$P(u_۱) = \frac{۱۰}{۲۵},$$

$$P(u_{۲}) = \frac{۹}{۲۵},$$
$$P(u_{۳}) = \frac{۶}{۲۵}.$$

۲.۳ (الف) زنجیر مارکوف از مرتبهٔ $k = ۲$ است، به قسمی که زنجیر می‌تواند خود را در $۴ = ۲^{۲} = m^k$ حالت مختلف بیابد. نمودار حالت در شکل (۸.۳) داده شده است.

(ب) احتمال چهار حالت $S_۱$، $S_۲$، $S_۳$ و $S_۴$ را می‌توان با کمک معادلات زیر محاسبه کرد:

$$P(۰۰) = P(۰۰) \,.\, P(۰|۰۰) + P(۱۰) \,.\, P(۰|۱۰),$$

$$P(۰۱) = P(۰۰) \,.\, P(۱|۰۰) + P(۱۰) \,.\, P(۱|۱۰),$$

$$P(۱۰) = P(۰۱) \,.\, P(۰|۰۱) + P(۱۱) \,.\, P(۰|۱۱),$$

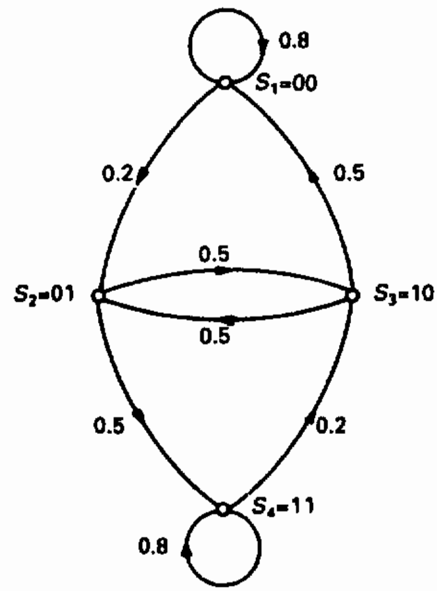
$$P(۱۱) = P(۱۱) \,.\, P(۱|۱۱) + P(۰۱) \,.\, P(۱|۰۱).$$

و

$$P(۰۰) + P(۰۱) + P(۱۰) + P(۱۱) = ۱$$

با قرار دادن احتمالهای مفروض در معادلات بالا نتیجه می‌شود

$$P(۰۰) = P(۰۰) \times ۰٫۸ + P(۱۰) \times ۰٫۵,$$



شکل ۸.۳- نمودار حالت تمرین (۲.۳)

$$P(۰۱) = P(۰۰) \times ۰٫۲ + P(۱۰) \times ۰٫۵,$$

$$P(۱۰) = P(۰۱) \times ۰٫۵ + P(۱۱) \times ۰٫۲,$$

$$P(۱۱) = P(۰۱) \times ۰٫۵ + P(۱۱) \times ۰٫۸,$$

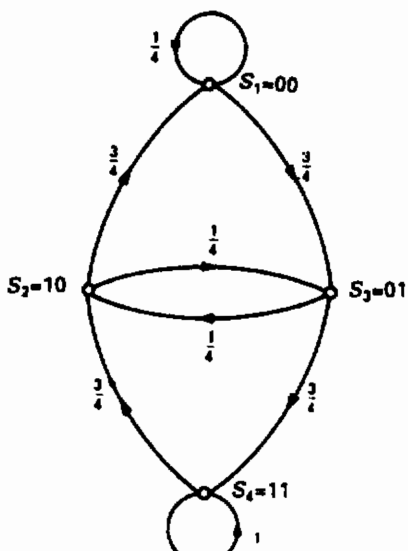
$$P(۰۰) + P(۰۱) + P(۱۰) + P(۱۱) = ۱.$$

با حلّ این چهار معادلهٔ سه مجهولی نتیجه می‌شود

$$P(۰۰) = P(۱۱) = \frac{۵}{۱۴},$$
$$P(۰۱) = P(۱۰) = \frac{۲}{۱۴}.$$

۳.۳ (الف) زنجیر مارکوف از مرتبهٔ $k = ۲$ است، به قسمی که زنجیر می‌تواند خود را در $۴ = ۲^۲ = m^k$ حالت که با ۰۰، ۰۱، ۱۰ و ۱۱ طرح شده است بیابد. $۸ = ۲^{۲+۱} = m^{k+1}$ انتقال و بنابراین ۸ احتمال انتقال وجود دارد. از اینها $۴ = ۲^{۲+۱} - ۲^۲ = m^{k+1} - m^k$ تا وجود دارند که می‌توان آزادانه انتخاب کرد. این احتمالها مفروض و بقیه ثابتند: $P(۱|۰۰) = ۱ - P(۰|۰۰) = \frac{۳}{۴}$ و تا آخر.

(ب) نمودار حالت فرمی مانند شکل (۹.۳) دارد.



شکل ۹.۳- نمودار حالت تمرین (۳.۳)

(پ) احتمال حالتها با رابطه‌های زیر داده شده‌اند

$$P(۰۰)=\frac{۱}{۴}P(۰۰)+\frac{۳}{۴}P(۱۰),$$

$$P(۰۱)=\frac{۳}{۴}P(۰۰)+\frac{۱}{۴}P(۱۰),$$

$$P(۱۰)=\frac{۱}{۴}P(۰۱)+\frac{۳}{۴}P(۱۱),$$

$$P(۱۱)=\frac{۳}{۴}P(۰۱)+\frac{۱}{۴}P(۱۱).$$

و

$$P(۰۰)+P(۱۰)+P(۰۱)+P(۱۱)=۱.$$

بنابراین داریم

$$P(۰۰)=P(۰۱)=P(۱۰)=P(۱۱)=\frac{۱}{۴}.$$

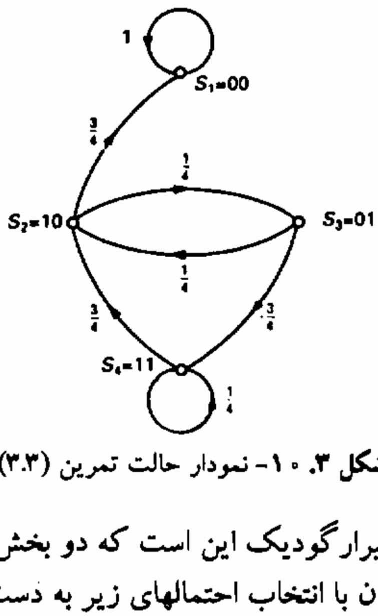
(ت) اکنون احتمال صفر را پیدا می‌کنیم

$$P(۰)=P(۰۰).P(۰|۰۰)+P(۰۱).P(۰|۰۱)+P(۱۰).P(۰|۱۰)+P(۱۱).P(۰|۱۱)$$

$$=\frac{۱}{۴}\times\frac{۱}{۴}+\frac{۱}{۴}\times\frac{۱}{۴}+\frac{۱}{۴}\times\frac{۳}{۴}+\frac{۱}{۴}\times\frac{۳}{۴}=\frac{۱}{۲},$$

و همچنین

$$P(۱)=\frac{۱}{۲}.$$



شکل ۳. ۱۰- نمودار حالت تمرین (۳.۳)

(ث) یک راه ساختن زنجیر غیرارگودیک این است که دو بخش جدا از هم بــه وجــود آوریم. این کار را می‌توان با انتخاب احتمالهای زیر به دست آورد

$$P(۰|۰۰)=۱,$$

یا

$$P(۰|۱۱)=۰.$$

(به قسمی که $P(۱|۰۰)=۰$ و $P(۱|۱۱)=۱$). با انتخاب $P(۰|۰۰)=۱$ زنجیر غیرارگودیک شکل (۳. ۱۰) نتیجه می‌شود.

۴.۳ (الف) نمودار حالت به صورتی است که در شکل (۱۱.۳) نشان داده شده است.

(ب) احتمالهای $u_۱$، $u_۲$ و $u_۳$ را از معادلات زیر می‌توان یافت

$$P(u_۱)=P(u_۱).P(u_۱|u_۱)+P(u_۲).P(u_۱|u_۲)+P(u_۳).P(u_۱|u_۳),$$

$$P(u_۲)=P(u_۱).P(u_۲|u_۱)+P(u_۲).P(u_۲|u_۲)+P(u_۳).P(u_۲|u_۳),$$

$$P(u_۳)=P(u_۱).P(u_۳|u_۱)+P(u_۲).P(u_۳|u_۲)+P(u_۳).P(u_۳|u_۳).$$

و شرط

$$P(u_۱)+P(u_۲)+P(u_۳)=۱.$$

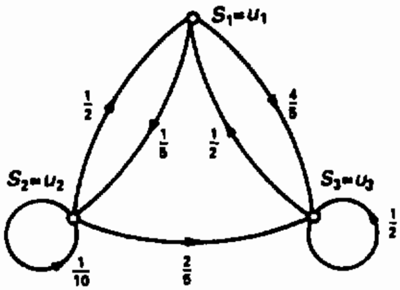
با به کارگیری مقدار داده شده نتیجه می‌شود

$$P(u_۱)=۰P(u_۱)+\frac{۱}{۲}P(u_۲)+\frac{۱}{۲}P(u_۳),$$

$$P(u_۲)=\frac{۱}{۵}P(u_۱)+\frac{۱}{۱۰}P(u_۲)+۰P(u_۳),$$

$$P(u_۳)=\frac{۴}{۵}P(u_۱)+\frac{۲}{۵}P(u_۲)+\frac{۱}{۲}P(u_۳),$$

$$P(u_۱)+P(u_۲)+P(u_۳)=۱.$$



شکل ۱۱.۳- نمودار حالت تمرین (۴.۳)

از این چهار معادلهٔ سه مجهولی نتیجه می‌شود که

$$P(u_۱)=\frac{۱}{۳},\qquad P(u_۲)=\frac{۲}{۲۷},\qquad P(u_۳)=\frac{۱۶}{۲۷}.$$

(پ) مقدار اطلاع منبع باحافظه با احتمالهای بالا برای نمادها به صورت زیر به دست می‌آید:

$$H(U)=-\sum_{i=۱}^{۳}P(u_i)\log P(u_i)$$
$$=-\frac{۱}{۳}\log\frac{۱}{۳}-\frac{۲}{۲۷}\log\frac{۲}{۲۷}-\frac{۱۶}{۲۷}\log\frac{۱۶}{۲۷}=۱٫۲۵\ \text{بیت / نماد}.$$

(ت) برای مقدار اطلاع منبع باحافظه برای یک انتقال دلخواه بنا بر تعریف داریم:

$$H(U_۲|U_۱)=-\sum_{i=۱}^{۳}\sum_{j=۱}^{۳}P(u_i)\,.\,P(u_j|u_i)\log P(u_j|u_i)$$
$$=\sum_{i=۱}^{۳}P(u_i)\left[-\sum_{j=۱}^{۳}P(u_j|u_i)\log P(u_j|u_i)\right].$$

با جایگزین کردن احتمالهای داده شده به دست می‌آوریم

$$H(U_۲|U_۱)=\frac{۱}{۳}\left[۰-\frac{۱}{۵}\log\frac{۱}{۵}-\frac{۴}{۵}\log\frac{۴}{۵}\right]$$
$$+\frac{۲}{۲۷}\left[-\frac{۱}{۲}\log\frac{۱}{۲}-\frac{۱}{۱۰}\log\frac{۱}{۱۰}-\frac{۲}{۵}\log\frac{۲}{۵}\right]$$
$$+\frac{۱۶}{۲۷}\left[-\frac{۱}{۲}\log\frac{۱}{۲}-۰-\frac{۱}{۲}\log\frac{۱}{۲}\right]$$
$$=۰٫۹۳\ \text{بیت / نماد}.$$

(ث) از رابطهٔ $H(U_۱,U_۲)=H(U_۱)+H(U_۲|U_۱)$ و این حقیقت که ممکن است فرض کنیم که منبع ماناست، یعنی

$$H(U_۱)=H(U_۲)=H(U),$$

در این صورت نتیجه می‌شود

$$H(U_۱,U_۲)=H(U_۱)+H(U_۲|U_۱)=۱٫۲۵+۰٫۹۳=۲٫۱۸\ \text{بیت / پیام}.$$

(ج) حشو به صورت زیر تعریف می‌شود

$$red = ۱ - \frac{H(U)}{\max\limits_{u} H(U)} = ۱ - \frac{H(U)}{\log n}.$$

با استفاده از مقادیر داده شده داریم

$$red = ۱ - \frac{۱٫۲۵}{\log ۳} = ۱ - \frac{۱٫۲۵}{۱٫۵۸} = ۰٫۲۱.$$

برای حشو مستقل داریم

$$red_{\infty} = ۱ - \frac{H_{\infty}(U)}{H(U)} = ۱ - \frac{H(U_۲ | U_۱)}{H(U)} = ۱ - \frac{۰٫۹۳}{۱٫۲۵} = ۰٫۲۶.$$

حشو کل ترکیبی از هر دو اندازه است و عبارت است از

$$red'_{\text{کل}} = ۱ - \frac{H_{\infty}(U)}{\max\limits_{u} H(U)} = ۱ - \frac{۰٫۹۳}{۱٫۵۸} = ۰٫۴۱.$$

(ج) از $M_{\max} = ۲^{l.\max H(U)}$ نتیجه می‌شود

$$\max\limits_{u} H(U) = \frac{۱}{l} \log M_{\max},$$

و از $M_{\infty} = ۲^{l.H_{\infty}(U)}$ نتیجه می‌شود

$$H_{\infty}(U) = \frac{۱}{l} \log M_{\infty}.$$

با جایگزین کردن عبارات به دست آمده برای $\max H(U)$ و $H_{\infty}(U)$ در عبارت حشو کل نتیجه می‌شود:

$$red_{\text{کل}} = ۱ - \frac{H_{\infty}(U)}{\max\limits_{u} H(U)} = ۱ - \frac{\log M_{\infty}}{\log M_{\max}}.$$

۵.۳ (الف) نمودار حالت فرمی مانند شکل (۱۲.۳) دارد.

(ب) احتمالهای $u_۱$، $u_۲$ و $u_۳$ از معادلات زیر به دست می‌آید:

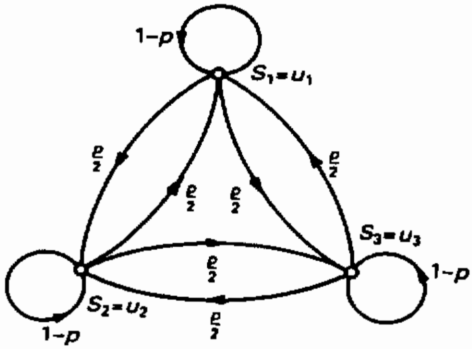
$$P(u_۱) = P(u_۱).P(u_۱|u_۱) + P(u_۲).P(u_۱|u_۲) + P(u_۳).P(u_۱|u_۳),$$

$$P(u_۲) = P(u_۱).P(u_۲|u_۱) + P(u_۲).P(u_۲|u_۲) + P(u_۳).P(u_۲|u_۳),$$

$$P(u_۳) = P(u_۱).P(u_۳|u_۱) + P(u_۲).P(u_۳|u_۲) + P(u_۳).P(u_۳|u_۳).$$

و

$$P(u_۱) + P(u_۲) + P(u_۳) = ۱.$$



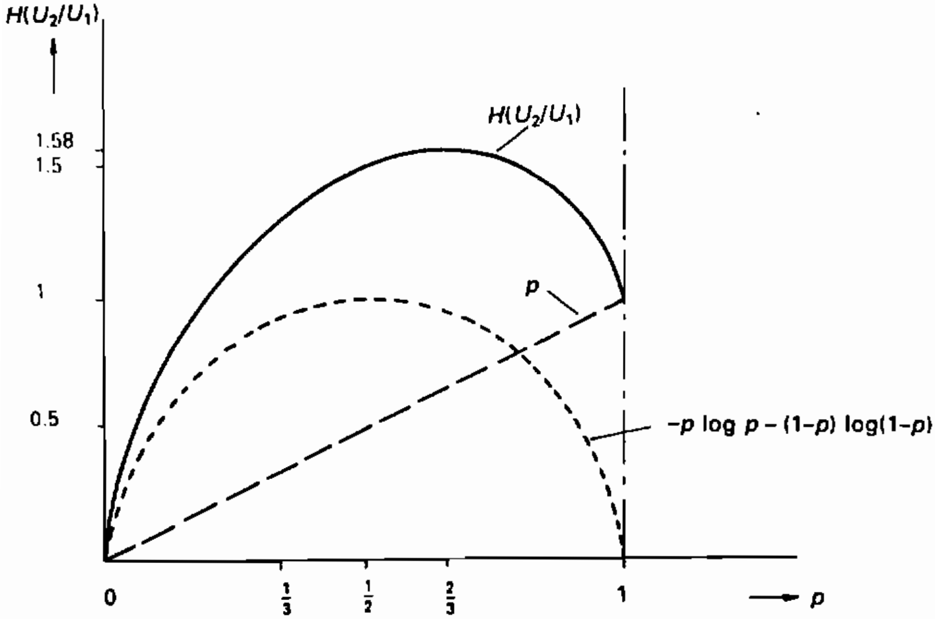
شکل ۱۲.۳- نمودار حالت تمرین (۵.۳)

با به کار بردن مقادیر داده شده نتیجه می‌شود که

$$P(u_۱) = P(u_۲) = P(u_۳) = \frac{۱}{۳}.$$

تقارن نمودار حالت، اشاره بر این دارد که هر حالت با احتمال مساوی رخ خواهد داد؛ بنابراین نتیجهٔ محاسبه شده با انتظارات، سازگار می‌باشد.

(پ) برای مقدار اطلاع نسبت به یک انتقال دلخواه داریم



شکل ۱۳.۳- مقدار اطلاع برای انتقال دلخواه به صورت تابعی از p

$$H(U_۲|U_۱) = -\sum_{i=۱}^{۳}\sum_{j=۱}^{۳} P(u_i)P(u_j|u_i)\log P(u_j|u_i)$$

$$= \sum_{i=۱}^{۳} P(u_i)\left[-\sum_{j=۱}^{۳} P(u_j|u_i)\log P(u_j|u_i)\right].$$

با جایگزین کردن مقادیر داده شده نتیجه می‌شود

$$H(U_۲|U_۱) = ۳\times\frac{۱}{۳}\left[-۲\frac{p}{۲}\log\frac{p}{۲}-(۱-p)\log(۱-p)\right]$$

$= p - p\log p - (۱-p)\log(۱-p)$ نماد / بیت.

شکل (۱۳.۳) را ببینید.

(ت) $H(U_۲|U_۱)$ ماکسیمم می‌شود اگر

$$\frac{dH(U_۲|U_۱)}{dp} = ۰.$$

یعنی

$$۱-\frac{۱}{\ln ۲}-\log p+\frac{۱}{\ln ۲}+\log(۱-p) = ۰.$$

از این معادله $p=\frac{۲}{۳}$ به دست می‌آید، و از آن نتیجه می‌شود که نماد / بیت $\max_u H(U) = \log ۳ = ۱{,}۵۸$ است.

(ث) اگر $p=۰$ زنجیر در همان حالت باقی بماند، دیگر هیچ عدم‌حتمیتی وجود ندارد؛ از این رو مقدار اطلاع برابر صفر است. اگر $p=۱$ زنجیر از هر حالت داده شده با شانس مساوی به یکی از دو حالت دیگر می‌رود؛ از این رو هر نماد یک بیت اطلاع می‌دهد. برای $p=\frac{۲}{۳}$ (ماکسیمم مقدار $H(U_۲|U_۱)$) هر انتقال دارای احتمال رخ‌داد یکسانی است. بنابراین رفتار زنجیر مانند حالتی است که سه نماد مستقلّند. اگر نماد قبلی معلوم باشد، عدم‌حتمیت کاهش نمی‌یابد. از این رو

$H(U_۲|U_۱) = \log n = \log ۳$ نماد / بیت.

۶.۳ (الف) برای محاسبهٔ مقدار اطلاع برای سه‌تایی‌ها ابتدا احتمالهای سه‌تایی را تعیین می‌کنیم. اینها عبارتند از

$$P(۰۰۰) = \frac{۵}{۱۴}\times\frac{۸}{۱۰} = \frac{۴}{۱۴},$$

$$P(۰۰۱)=\frac{۵}{۱۴}\times\frac{۲}{۱۰}=\frac{۱}{۱۴},$$

$$P(۰۱۰)=\frac{۱}{۷}\times\frac{۱}{۲}=\frac{۱}{۱۴},$$

$$P(۰۱۱)=\frac{۱}{۷}\times\frac{۱}{۲}=\frac{۱}{۱۴},$$

$$P(۱۰۰)=\frac{۱}{۷}\times\frac{۱}{۲}=\frac{۱}{۱۴},$$

$$P(۱۰۱)=\frac{۱}{۷}\times\frac{۱}{۲}=\frac{۱}{۱۴},$$

$$P(۱۱۰)=\frac{۵}{۱۴}\times\frac{۲}{۱۰}=\frac{۱}{۱۴},$$

$$P(۱۱۱)=\frac{۵}{۱۴}\times\frac{۸}{۱۰}=\frac{۴}{۱۴}.$$

از این رو برای (سه‌تایی)H نتیجه می‌شود

$$H(\text{سه‌تایی})=-۲\times\frac{۴}{۱۴}\log\frac{۴}{۱۴}-۶\times\frac{۱}{۱۴}\log\frac{۱}{۱۴}=۲٫۶۷\ \text{بیت / سه‌تایی}.$$

در این رابطه مقدار اطلاع بر نماد به دست می‌آید.

$$H_۳(U)=\frac{۱}{۳}H(\text{سه‌تایی})=۰٫۸۹\ \text{بیت / نماد}.$$

(ب) احتمالهای دوتایی مساوی با احتمالهای حالتها هستند به قسمی که:

$$H(\text{دوتایی})=-۲\times\frac{۵}{۱۴}\log\frac{۵}{۱۴}-۲\times\frac{۱}{۷}\log\frac{۱}{۷}=۱٫۸۶\ \text{بیت / دوتایی},$$

و

$$H_۲(U)=\frac{۱}{۲}H(\text{دوتایی})=۰٫۹۳\ \text{بیت / نماد}.$$

(پ) اکنون مقدار $H_۱(U)$ از احتمالهای $P(۰)$ و $P(۱)$ محاسبه می‌شود. این احتمالها را می‌توان از رابطهٔ زیر به دست آورد

$$P(۰)=\sum_{i=۱}P_i(\text{دوتایی})\frac{n_{۰,j}}{n_{۰,j}+n_{۱,j}},$$

که در آن $n_{۰,j}$ تعداد صفرها در دوتایی iام و $n_{۱,j}$ تعداد یکها در دوتایی iام می‌باشد.

یادآوری: این فرمول تنها برای کدواژه‌های با طول مساوی معتبر می‌باشد. بنابراین

$$P(۰) = \frac{۵}{۱۴} \times ۱ + \frac{۵}{۱۴} \times ۰ + \frac{۱}{۷} \times \frac{۱}{۲} + \frac{۱}{۷} \times \frac{۱}{۲} = \frac{۱}{۲},$$

و

$$P(۱) = \frac{۱}{۲}.$$

از این به دست می‌آید

$$H_۱(U) = -\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۲}\log\frac{۱}{۲} = \text{۱ بیت / نماد}.$$

(ت) برای $F_N(U)$ می‌توان به دست آورد که (بخش (۲.۳) را ببینید)

$$N\, H_N(U) = (N-۱) H_{N-۱}(U) + F_N(U).$$

با به کار بردن مقادیر معلوم $H_N(U)$ در این صورت نتیجه می‌شود

$$F_۱(U) = H_۱(U) = \text{۱ بیت / نماد},$$

$$F_۲(U) = ۲H_۲(U) - H_۱(U) = \text{۰٫۸۶ بیت / نماد},$$

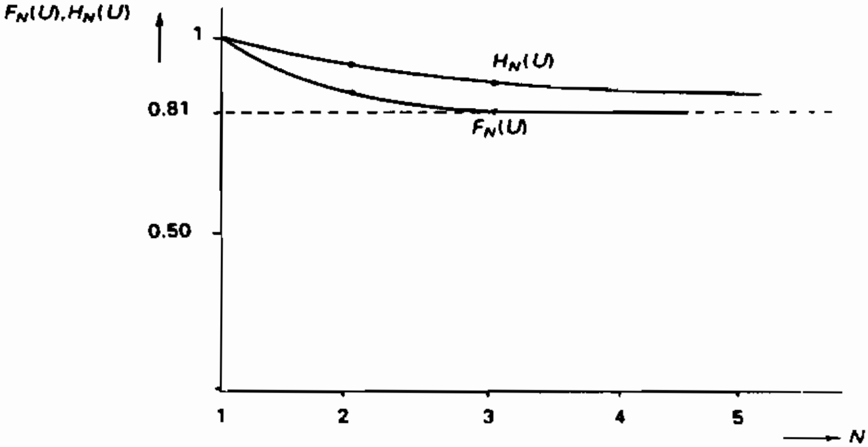
$$F_۳(U) = ۳H_۳(U) - ۲H_۲(U) = \text{۰٫۸۱ بیت / نماد}.$$

(ث) با افزایش مقدار N، $F_N(U)$ با احتمالهای انتقالی که مقدار آن بستگی بـه افزایـش تعداد نمادهای قبلی دارد تعیین خواهد شد. بنابراین عدم‌حتمیـت کـاهش می‌یـابد، همچنان که $F_N(U)$ نیز کاهش خواهد یافت.

(ج) مقدار $F_۴(U)$ برابر $F_۳(U)$ است. دلیل آن این است که زنجیر مـارکوف از مرتبـهٔ دوم می‌باشد. اکنون دو نماد قبلی بر احتمالهای انتقال تأثیر دارند، بقیهٔ نمادها هیـچ اثری ندارند. این بدین معناست که مقدار $F_N(U)$ دیگر بـرای $N \geq ۳$ هیـچ تغییـر نمی‌کند. مقدار $H_۴(U)$ از مقدار $H_۳(U)$ کمتر است، زیرا با محاسـبهٔ $H_N(U)$ دو نماد اوّل پیامها بدون تأثیر باقی می‌مانند. با وجود این، روشن است کـه ایـن اثـر کاهش می‌یابد پس $H_N(U)$ یک تابع کاهشی یکنواخت از N با مقدار حدّی زیـر می‌باشد:

$$\lim_{N\to\infty} H_N(U) = F_۳(U).$$

(چ) طرح شکل (۱۴.۳) مسیر $F_N(u)$ و $H_N(u)$ را به صورت تابعی از N می‌دهد.

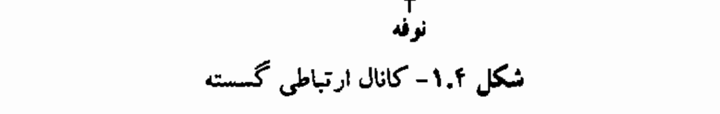


شکل ۱۴.۳- $F_N(U)$ و $H_N(U)$ به عنوان تابعی از N

۴
کانال ارتباطی گسسته

۱.۴ ظرفیت کانالهای بدون نوفه

مقدار اطلاع حاصل از یک منبع اطلاع را می‌توان برحسب بیت بر نماد بیان کرد، همان‌گونه که در فصلهای پیشین اغلب مطرح بود. با وجود این، اگر این امر که یک منبع اطلاع در مدّت زمان معیّنی دنباله‌ای از نمادها را تولید می‌کند، حائز اهمیت باشد، بیان مقدار اطلاع برحسب بیت بر ثانیه می‌تواند مفید واقع شود. البته رابطهٔ بین بیت بر نماد از یک طرف و بیت بر ثانیه از طرف دیگر با متوسط مدّت نمادها تعیین می‌گردد. وقتی که مقدار اطلاع تولید شده توسط یک منبع اطلاع باید به مقصدی فرستاده شود، استفاده از یک کانال ارتباطی که قادر به انتقال اطلاع معرّفی شده است ضروری است. برای کانال ارتباطی در این فصل الگوی داده شده در شکل (۱.۴) را که مشابه الگوی داده شده در شکل (۸.۱) است در نظر خواهیم گرفت.



شکل ۱.۴- کانال ارتباطی گسسته

معمولاً اطلاع حاصل از منبع به گونه‌ای انتقال می‌یابد که مستقیماً نمی‌تواند از کانال عبور کند. در این صورت اطلاع باید مناسب با کانال رمزگذاری شود. علاوه بر این، در

فرایند کدگذاری کاهش حشو، برای این که مقدار اطلاعی که باید از طریق کانال ارسال شود تا حدّ ممکن محدود گردد، مطلوب است، که کار مشکلی است و در فصلهای قبل مطالعه شده است. نمادهای به دست آمده نسبت به کانال به سیگنالی که سازگار با ویژگیهای فیزیکی کانال است تبدیل می‌شوند که ممکن است هنگام انتقال در اثر نوفه تغییر شکل دهند. در نتیجه یک کدنماد و یا پیام ارسال شده مانند آنچه که پس از عبور از کانال کدگشایی می‌شود نیست؛ زیرا ممکن است نمادها در نوفه تغییر یافته باشند. از دیدگاه نظریهٔ اطلاع روشن است که به احتمال خطایی که در حین انتقال یک نماد رخ می‌دهد علاقه‌مند باشیم.

علاوه بر این نکته، جنبهٔ دیگری وجود دارد که ارزش توجّه دارد. از نظر ویژگیهای فیزیکی کانال مقدار اطلاعی که منتقل می‌شود محدود خواهد بود. بنابراین شخص نظریه‌پرداز اطلاع، به ***ظرفیت*** الگوی ارتباطی نیز علاقه‌مند است. این ظرفیت ماکسیمم مقدار اطلاعی است که می‌تواند از طریق کانال داده شده عبور کند.

تعریف ۱.۴

ظرفیت C یک کانال بدون نوفه گسسته به صورت زیر تعریف می‌شود:

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T} \text{ ثانیه / بیت}, \qquad (۱.۴)$$

که در آن $N(T)$ تعداد پیامهای مجاز با مدّت T می‌باشد. ●

از این تعریف نتیجه می‌شود که برای T بزرگ می‌توان در حدود 2^{CT} پیام مختلف را از طریق کانال در مدّت زمان T انتقال داد.

در ادامه ظرفیت کانالهای بدون نوفه گسسته که نمادها همگی دارای مدّت یا طول مساوی نیستند ارائه خواهد شد. برای مثال کد مُرس را مقایسه کنید که در آن نقطه دارای مدّت کوتاهتری از خط تیره است.

همان‌طور که در فصل سوم دیدیم ساختمان کدواژه‌ها به صورت دنباله‌ای از نمادها را می‌توان به عنوان نتیجه‌ای از یک فرایند مارکوف در نظر گرفت. هر نماد یک پیام را می‌توان به صورت وابسته به نمادهای قبلی آن در نظر گرفت. حالتهای فرایند مارکوف نماد بعدی را تعیین می‌کنند.

فرض کنید $N_j(T)$ تعداد پیامهای مختلف ممکن با مدّت T باشد که در حالت S_j ختم می‌شود. فرض کنید t_{ij}^{s} مدّت نماد s باشد که به موجب آن فرایند از حالت S_i به S_j

می‌رود. در این صورت $N_i(T-t_{ij}^s)$ تعداد پیامهایی است که با انتقال از حالت S_i به حالت S_j با نماد s ختم می‌شود. واضح است که با جمع‌بندی همهٔ حالتهای S_i و تمام نمادهای s که موجب انتقال به حالت S_j می‌شوند، تعداد $N_j(T)$ به دست می‌آید:

$$N_j(T)=\sum_i\sum_s N_i(T-t_{ij}^s). \qquad (۲.۴)$$

با جمع‌بندی روی تمام S_jها، $N(T)$ به دست می‌آید

$$N(T)=\sum_j N_j(T). \qquad (۳.۴)$$

با یافتن عبارتی برای $N_j(T)$ و از این رو برای $N(T)$، ظرفیت را می‌توان پیدا کرد.

قضیه ۱.۴

فرض کنید A ماتریسی با درایه‌های زیر باشد

$$a_{ij}=\sum_s X^{-t_{ij}^s}-\delta_{ij}, \qquad (۴.۴)$$

که δ_{ij} نماد کرونکر است ($\delta_{ij}=1$ برای $i=j$، در غیر این صورت $\delta_{ij}=0$) و t_{ij}^s مدّت نماد s است که به موجب آن فرایند از حالت S_i به S_j می‌رود.

ظرفیت کانال بدون نوفه گسسته با نمادهایی با مدّت نامساوی به صورت زیر داده می‌شود

$$C=\log X_0, \qquad (۵.۴)$$

که در آن x_0 بزرگترین X مثبت است که برای آن دترمینان ماتریس A برابر صفر است

$$|A|=0.$$

برهان

در واقع معادلهٔ زیر را حل می‌کنیم

$$N_j(T)=\sum_i\{\sum_s N_i(T-t_{ij}^s)\} \qquad (۶.۴)$$

زیرا در این صورت می‌توان $N(T)$ و به دنبال آن ظرفیت را با کمک تعریف (۱.۴) محاسبه کرد.

در حقیقت، معادلهٔ (۶.۴) یک معادلهٔ تفاضلی خطّی است. برای حلّ معادلات تفاضلی خطّی روشهایی به دست آورده‌اند که مشابه روشهای حلّ معادلات دیفرانسیل خطّی

می‌باشند. فرض کنید که رابطهٔ بین تابع $X(n)$ و $X(n-1)$ به صورت زیر باشد

$$X(n)-\alpha X(n-1)=0.$$

در این صورت یک جواب خاص آن به صورت زیر است

$$X(n)=c\lambda^n.$$

با قرار دادن این جواب در معادلهٔ تفاضلی به دست می‌آوریم:

$$c\lambda^n-\alpha c\lambda^{n-1}=0.$$

و با حلّ این معادلهٔ مشخّص نتیجه می‌شود

$$\lambda=\alpha,$$

و

$$X(n)=c\alpha^n.$$

برای معادلات تفاضلی خطّی از مرتبهٔ بالاتر با روشی مشابه به دست می‌آوریم

$$X(n)-\alpha_1 X(n-1)-\alpha_2 X(n-2)-\cdots-\alpha_p X(n-p)=0.$$

اکنون p جواب خاص به صورت $c_k\lambda_k^n$ داریم و وقتی آنها را با یکدیگر جمع کنیم جواب معادلهٔ تفاضلی به دست می‌آید

$$X(n)=c_1\lambda_1^n+c_2\lambda_2^n+\cdots+c_p\lambda_p^n.$$

با قرار دادن هر جواب خاص در معادله، معادلهٔ مشخّصی به دست می‌آید، برای مثال:

$$c_k\lambda_k^n-\alpha_1 c_k\lambda_k^{n-1}-\alpha_2 c_k\lambda_k^{n-2}-\cdots-\alpha_p c_k\lambda_k^{n-p}=0$$

که می‌توان λ_k را تعیین کرد.

اگر $\lambda_1>\lambda_2>\ldots>\lambda_p$ و $n\to\infty$، در این صورت تقریباً داریم: $X(n)=c_1\lambda_1^n$.

تذکّر: مسلّماً جملات دیگری که دارای خاصیت نوسانی هستند باید برای T متناهی اضافه کرد.

اکنون حل معادلهٔ (۶.۴) به شکل زیر است

$$N_j(T)=\alpha_j X^T. \tag{۷.۴}$$

با جایگزینی این جواب به دست می‌آوریم

$$\alpha_j X^T = \sum_i \sum_s \alpha_i X^{T - t_{ij}^s} . \tag{۸.۴}$$

با تقسیم بر X^T نتیجه می‌شود

$$\sum_i \sum_s \alpha_i X^{-t_{ij}^s} - \alpha_j = ۰. \tag{۹.۴}$$

α_j را به صورت زیر می‌نویسیم

$$\alpha_j = \sum_i \delta_{ij} \alpha_i ,$$

با

$$\delta_{ij} = ۱ \quad , \quad i = j ,$$

و

$$\delta_{ij} = ۰ \quad , \quad i \neq j .$$

در این صورت به دست می‌آوریم

$$\sum_i \alpha_i \{ \sum_s X^{-t_{ij}^s} - \delta_{ij} \} = ۰. \tag{۱۰.۴}$$

i معادلهٔ بین ابروها مستقلند. اگر این فرض را که همهٔ α_i صفر هستند مستثنا کنیم، در این صورت تساوی تنها وقتی می‌تواند برقرار باشد که دترمینان رونسکی[1] برابر صفر باشد. بنابراین

$$\begin{bmatrix} \sum_s X^{-t_{۱۱}^s} - ۱ & \sum_s X^{-t_{۱۲}^s} & \dots & \dots \\ \sum_s X^{-t_{۲۱}^s} & \sum_s X^{-t_{۲۲}^s} - ۱ & \dots & \dots \\ \dots & & & \\ \dots & & & \end{bmatrix} = ۰.$$

به شکل مختصر شده داریم:

$$|A| = ۰ \quad \text{با} \quad a_{ij} = \sum_s X^{-t_{ij}^s} - \delta_{ij} . \tag{۱۱.۴}$$

با بزرگترین ریشهٔ مثبت، $X_۰$، این تابع نتیجه می‌شود

$$N_j(T) = \alpha_j X_۰^T , \tag{۱۲.۴}$$

و بنابراین

$$N(T) = \sum_j \alpha_j X_۰^T . \tag{۱۳.۴}$$

1. Wronskian

از این رابطه برای ظرفیت کانال ارتباطی که از این سیستم کدگذاری استفاده می‌کند نتیجه می‌شود که

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T}$$

$$= \lim_{T\to\infty} \frac{\log(X_\circ^T \sum_j \alpha_j)}{T}$$

$$= \lim_{T\to\infty} \frac{T \log X_\circ}{T} + \lim_{T\to\infty} \left\{ \frac{\log \sum_s \alpha_j}{T} \right\}$$

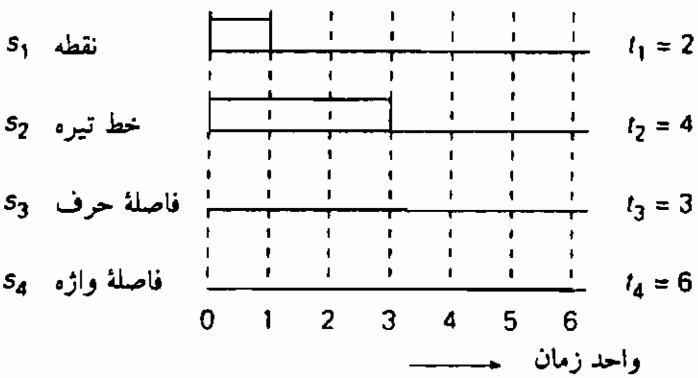
$$= \log X_\circ . \qquad (۱۴.۴)$$

■

در مثال زیر یک کد مرس را در نظر می‌گیریم و ظرفیت را محاسبه می‌کنیم.

مثال ۱.۴

کد مرس دارای چهار نماد؛ یک نقطه، یک خط تیره، یک فاصلهٔ حرف و یک فاصلهٔ واژه، هر یک با مدّت خودشان می‌باشد. شکل (۲.۴) را ببینید. یک شرط وجود دارد، یعنی این‌که یک فاصلهٔ حرف یا یک فاصلهٔ واژه همواره باید با یک نقطه یا یک خط تیره دنبال شود. بنابراین فاصله‌ها مجاز نیستند که یکدیگر را دنبال کنند. پس دو حالت می‌توان تشخیص داد: حالت $S۱$ بعد از یک فاصله و حالت $S۲$ پس از یک نقطه یا یک خط تیره. این نمودار حالت داده شده در شکل (۳.۴) را می‌دهد. برای درایه‌های ماتریس A به دست می‌آوریم



شکل ۲.۴- کد مُرس

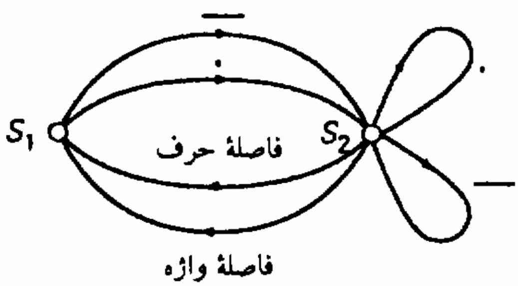
$$a_{11} = \sum_s X^{-t^s_{11}} - 1 = -1,$$
$$a_{12} = \sum_s X^{-t^s_{12}} \quad = X^{-t_1} + X^{-t_2} \quad = X^{-2} + X^{-4},$$
$$a_{21} = \sum_s X^{-t^s_{21}} \quad = X^{-t_3} + X^{-t_4} \quad = X^{-3} + X^{-6},$$
$$a_{22} = \sum_s X^{-t^s_{22}} - 1 = X^{-t_1} + X^{-t_2} - 1 = X^{-2} + X^{-4} - 1,$$

و از این رو دترمینان رونسکی منجر می‌شود به

$$|A| = \begin{bmatrix} -1 & X^{-2} + X^{-4} \\ X^{-3} + X^{-6} & X^{-2} + X^{-4} - 1 \end{bmatrix} = 0,$$

یا

$$f(X) = X^{-2} + X^{-4} + X^{-5} + X^{-7} + X^{-8} + X^{-10} - 1 = 0,$$



شکل ۳.۴- نمودار حالت نسبت به کد مُرس

تقریبی از بزرگترین ریشه مثبت X را به طور ترسیمی به دست می‌آوریم و سپس با چند محاسبه $f(X)$ حاصل می‌شود (شکل (۴.۴) را ببینید).

$$X_0 \approx 1{,}454.$$

بنابراین ظرفیت برابر است با

$C = \log X_0 = \log 1{,}454 = 0{,}54$ بیت / ثانیه.

اگر شرط را حذف کنیم، نمادها ممکن است یکدیگر را دنبال کنند و سیستم در یک حالت باقی بماند. اکنون نمودار در شکل (۵.۴) داده شده است.

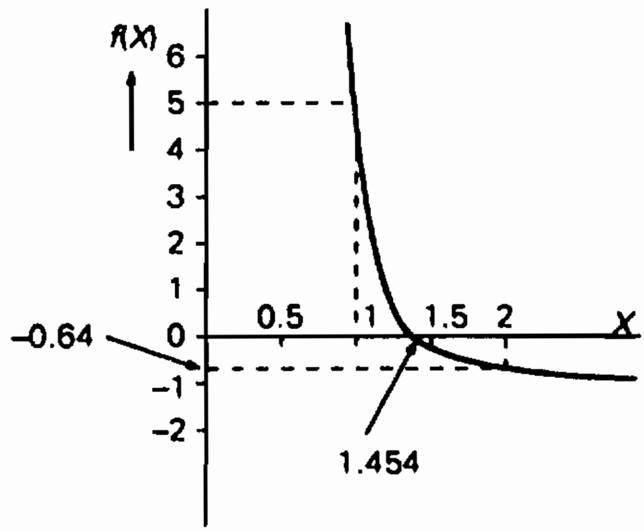
$$\begin{aligned}|A| = a_{11} &= \sum_s X^{-l_{11}^s} - 1\\ &= X^{-l_1} + X^{-l_2} + X^{-l_3} + X^{-l_4} - 1\\ &= X^{-2} + X^{-3} + X^{-4} + X^{-6} - 1 = 0.\end{aligned}$$

اکنون به طور ترسیمی همراه با چند محاسبه $f(X)$ به دست می‌آید:

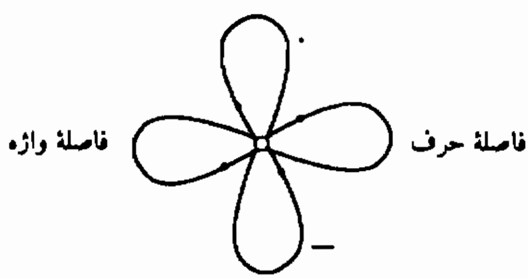
$$X_0 \approx 1{,}51.$$

حال ظرفیت برابر است با

$C = \log 1{,}51 = 0{,}59$ ثانیه / بیت



شکل ۴.۴- تقریب عددی ظرفیت با در نظر گرفتن $f(X) = 0$



شکل ۵.۴- نمودار حالت بدون شرط

بنابراین با کمک کد مرس با شرط و بدون شرط به ترتیب حداکثر ۵۴ر۰ و ۵۹ر۰بیت بر واحد زمان می‌توان انتقال داد. این بستگی به روشی که کد در آن به کار رفته است دارد و خواهی‌نخواهی این حدها به دست می‌آید. ▲

۲.۴ ظرفیت کانالهای نوفه‌دار

در کانالهای نوفه‌دار قبل از این‌که جنبه‌هایی مانند احتمال خطا و ظرفیت را مورد توجّه قرار دهیم، ابتدا توصیف کانال ارتباطی را با تفصیل بیشتری بررسی خواهیم کرد. فرض کنید، پس از کدگذاری و از این رو در ورودی منبع اطلاع (شکل (۱.۴) را ببینید)، نمادهای u_i حاصل از منبع اطلاع به کدنمادهای x_i متعلّق به الفبای $X=\{x_1,x_2,\ldots,x_m\}$ رمزگذاری شده‌اند، در حالی که پس از عبور از کانال نمادهای y_i متعلّق به الفبای $Y=\{y_1,y_2,\ldots,y_n\}$ دریافت می‌شوند. مقدار اطلاع در ورودی را می‌توان با $H(X)$ بیت بر ثانیه و در خروجی با $H(Y)$ بیت بر ثانیه نشان داد. اگر فرض کنیم که کدواژه $\tilde{x}_i=(x_{i1},x_{i2},\ldots,x_{iN})$ با طول L معرّفی شده است و به عنوان نتیجه‌ای از آن کدواژه $\tilde{y}_j=(y_{j1},y_{j2},\ldots,y_{jN})$ دریافت شده است، در این صورت احتمال $\tilde{y}_j$ به شرط $\tilde{x}_i$ را می‌توان به صورت زیر به دست آورد

$$q(\tilde{y}_j|\tilde{x}_i)=q(y_{j1},y_{j2},\ldots,y_{jL}|x_{i1},x_{i2},\ldots,x_{iL}). \qquad (۱۵.۴)$$

در موارد بسیاری می‌توان فرض کرد که کدنماد دریافت شده y_{jk} تنها به نماد ارسال شده x_{ik} بستگی دارد و به نمادهای قبلی $x_{i1},\ldots,x_{i,k-1}$ یا نمادهای قبلی $y_{j1},\ldots,y_{j,k-1}$ بستگی ندارد. در این صورت ***کانال بی‌حافظه*** است، در این حالت احتمال انتقال کدواژهٔ $\tilde{x}_i$ به کدواژهٔ $\tilde{y}_j$ را می‌توان به احتمالهای انتقال نمادهای تکی تقسیم کرد، یعنی

$$q(\tilde{y}_j|\tilde{x}_i)=\prod_{k=1}^{L}q(y_{jk}|x_{ik}). \qquad (۱۶.۴)$$

بنابراین وقتی فرض شده است که منبع بی‌حافظه باشد بازهم می‌توان ویژگیهای کانال گسسته را برای هر نماد ارسال شده به جای هر کدواژه که به طور قابل ملاحظه‌ای تحلیل را ساده می‌کند، تعیین نمود. اکنون می‌توان کانال گسسته را با ارائهٔ ماتریس احتمالهای شرطی $q(y_j|x_i)$ برای تمام نمادهای y_j به شرط x_i مشخّص کرد. با استفاده از علامت اختصاری برای احتمالهای شرطی، یعنی $q(y_j|x_i)=q_{ji}$ به اصطلاح ***ماتریس کانال*** به شکل

زیر در می‌آید:

$$Q=\begin{bmatrix} q_{11} & q_{21} & \cdots & q_{n1} \\ q_{12} & q_{22} & \cdots & q_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ q_{1m} & q_{2m} & \cdots & q_{nm} \end{bmatrix}. \qquad (۱۷.۴)$$

چون i نماد ورودی را نشان می‌دهد و هر نماد ورودی به عنوان نتیجه دارای یک نماد خروجی است، مجموع هر ردیف ماتریس باید برابر ۱ باشد. احتمال رخ‌داد نماد خروجی y_j را می‌توان به صورت زیر تعیین کرد

$$q(y_j)=\sum_{i=1}^{m} p(x_i)q_{ji},$$

در حالی که احتمالهای پسین x_i با به کار بردن قضیهٔ بیز به دست می‌آیند

$$p(x_i|y_j)=\frac{p(x_i)\cdot q(y_j|x_i)}{q(y_j)}. \qquad (۱۸.۴)$$

با بهره‌وری از این رابطه می‌توان مقدار اطلاع شرطی $H(X|Y)$ و $H(Y|X)$ را بنا به تعریف (۳.۱) محاسبه کرد

$$H(X|Y)=-\sum_{i=1}^{m}\sum_{j=1}^{n} q(y_j)\cdot p(x_i|y_j)\log p(x_i|y_j), \qquad (۱۹.۴)$$

و

$$H(Y|X)=-\sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i)\cdot q(y_j|x_i)\log q(y_j|x_i). \qquad (۲۰.۴)$$

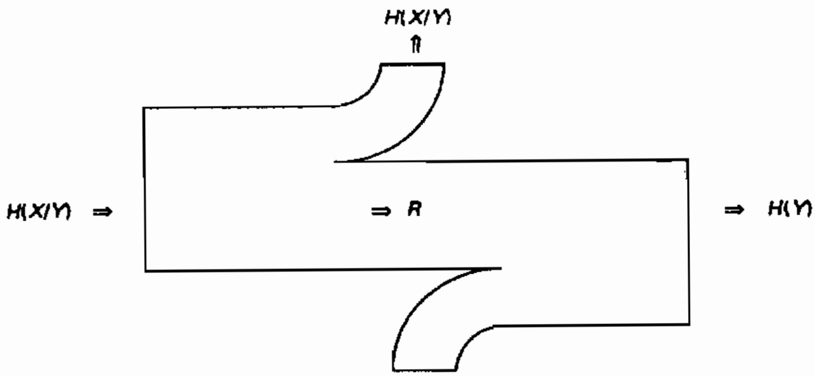
مقدار اطلاع شرطی $H(X|Y)$، به عبارت دیگر عدم‌حتمیت دربارهٔ x اگر y دریافت شده باشد را ***ابهام*** می‌نامند. مقدار اطلاع شرطی $H(Y|X)$ را می‌توان به عنوان عدم‌حتمیت در y اگر x معلوم باشد در نظر گرفت. این عدم‌حتمیتی است که با نوفه معرّفی شده است و ***بی‌ارتباطی*** نامیده می‌شود.

اگر علاوه بر $H(X|Y)$ و $H(Y|X)$ مقادیر اطلاع حاشیه‌ای $H(X)$ و $H(Y)$ نیز معلوم باشند، مقدار اطلاع متقابل را می‌توان براساس معادلهٔ (۳۶.۱) محاسبه کرد:

$$\begin{aligned} I(X;Y) &= H(Y)-H(Y|X) \\ &= H(X)-H(X|Y). \end{aligned} \qquad (۲۱.۴)$$

از بیان اخیر نتیجه می‌شود که اطلاع متقابل را می‌توان بــه صــورت عدم‌حتمیــت در طرف گیرنده دربارهٔ نماد ارسال شدهٔ x قبل از این‌که نمــاد y ارســال شــده باشــد منهــای عدم‌حتمیت باقی‌مانده پس از این‌که نماد ارسال شده باشد در نظر گرفت. یعنی می گوییم که $I(X;Y)$ وابسته به مقدار اطلاعی می‌باشد که به سوی دیگر کانال منتقل شده است. گــاهی اوقات نماد R نیز به جای $I(X;Y)$ به کار برده می‌شود و در این صورت آن را ***نرخ ارسال*** می‌نامیم. بدون کاستی در کلیّت مسأله می‌توان آن را برحسب بیت بــر ثانیــه بیــان کــرد. بدیهی است برای کانالی که تغییر شکل نیافته اســت داریــم: $H(X|Y) = H(Y|X) = 0$. در این حالت $I(X;Y)$ مقدار ماکسیمم $H(X)$ را می‌گیرد. انتقال اطلاع را می‌توان به صـــورت نموداری با شیوهٔ نشان داده شده در شکل (۶.۴) ترسیم کرد.

توجّه کنید که تنها احتمالهای انتقال $q(y_j|x_i)$ برای کانال داده شده ثابتند. از این رو این احتمالهای انتقال اثر نوفه را همان‌طور که بوده است نشــان می‌دهنـد. بـا وجـود ایـن، $I(X;Y)$ تنها به احتمالهای $q(y_j|x_i)$ وابسته نیست، بلکه بــه احتمالهــای $p(x_i)$ نمادهــای ورودی نیز وابسته است، معادلهٔ (۲۱.۴) را ببینید، بنابراین هنوز هم می‌توانــد بــرای کانــال داده شده با تغییر احتمالهای $p(x_i)$ تغییر کند.



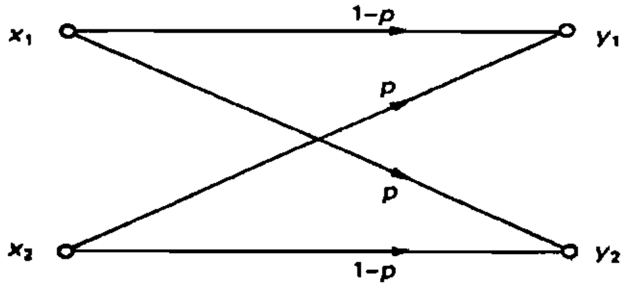
شکل ۶.۴- طرح برای انتقال اطلاع

ساده‌ترین کانال، کانالی است که در آن پیامهایی که ارائه شده‌اند با الفبای دودویـــی $X = \{x_1, x_2\}$ رمزگذاری شده‌اند و پیامهایی با الفبای دودویی $Y = \{y_1, y_2\}$ نتیجه می‌شود، وقتی x_1 و x_2 ارسال می‌شوند احتمالهای دریافت y_1 و y_2 بـــه ترتیـب هـر دو برابـر $1-p$ می‌باشد. چنین کانالی را ***کانال متقارن دودویی***، BSC، می‌نامند. ماتریس کانال برای چنین

کانالی عبارت است از

$$Q=\begin{bmatrix} ۱-p & p \\ p & ۱-p \end{bmatrix}.$$

با نوشتن نمادهای ورودی در زیر یکدیگر در سمت چپ و نمادهای خروجی در سمت راست و با نشان دادن انتقالها با احتمالهایشان با پیکان می‌توان این کانال را به صورت شکل (۷.۴) نمایش داد.



شکل ۷.۴- کانال متقارن دودویی

مثال ۲.۴

مقدار اطلاع ارسال شده $I(X;Y)$ یک کانال متقارن دودویی را به آسانی می‌توان محاسبه کرد. فرض کنید $p(x_۱)=\alpha$ و بنابراین $p(x_۲)=۱-\alpha$، و با فرض $q(y_۱)=\beta$ داریم $q(y_۲)=۱-\beta$. در این صورت داریم:

$$\begin{aligned} \mathsf{R} &= H(Y)-H(Y|X) \\ &= -\beta\log\beta-(۱-\beta)\log(۱-\beta)+p\log p+(۱-p)\log(۱-p) \end{aligned}$$

با

$$\beta=\alpha(۱-p)+(۱-\alpha)p. \qquad \blacktriangle$$

علاوه بر کانالهای متقارن دودویی همچنین می‌توان کانالهای غیر- دودویی و نامتقارن را تشخیص داد. چنین کانالهایی را در بخش زیر بررسی خواهیم کرد.

در مطالب گذشته دیدیم که از نقطه‌نظر دریافت کننده عدم‌حتمیت قبل از این که پیام دریافت شود برابر با $H(X)$ است. پس از این که گیرنده یک پیام یا نماد را دریافت کرد عدم‌حتمیت دربارهٔ نماد یا پیام ارسال شده به $H(X|Y)$ کاهش می‌یابد.

از این رو در واقع مقدار اطلاعی برابر با $\mathsf{R}=H(X)-H(X|Y)$ بر کانال انتقال داده

شده است. اکنون ظرفیت C یک کانال نوفه‌دار به صورت ماکسیمم مقدار اطلاعی که از کانال می‌توان انتقال داد تعریف می‌شود. برای ختم این مبحث همهٔ منابع اطلاع ممکن را به کانال وصل می‌کنیم به قسمی که از همهٔ توزیعهای احتمال ممکن $p(x_i)$ برای $i = ۱,\ldots,m$ استفاده شود.

تعریف ۲.۴

ظرفیت کانال نوفه‌دار گسسته به صورت زیر تعریف می‌شود

$$C = \max_{p(x)} R = \max_{p(x)} I(X;Y) = \max_{p(x)} \{H(Y) - H(Y|X)\}$$

$$= \max_{p(x)} \{H(X) - H(X|Y)\} \text{ نماد / بیت}. \qquad (۲۲.۴)$$

اگر کانال تغییر نیابد، یعنی $H(X|Y) = H(Y|X) = ۰$، C مقدار ماکسیمم $H(X)$ می‌باشد. چون $\max_{p(x)} H(X) = \log m$ در این صورت ظرفیت برابر است با

$$C = \log m \text{ نماد / بیت} \qquad (۲۳.۴)$$

اگر فرض کنیم که نمادها دارای مدّت مشترک t ثانیه باشند در این صورت ظرفیت کانال بر ثانیه به صورت زیر است

$$C = \frac{۱}{t} \log m \text{ ثانیه / بیت}. \qquad (۲۴.۴)$$

که متناظر با تعریف (۱.۴) است و می‌توان آن را به صورت زیر دید. اگر کدواژه‌های با طول L و با مدّت T شامل نمادهایی از یک الفبا با طول m را در نظر بگیریم در این صورت تعداد کدواژه‌های ممکن برابر است با: $N(T) = m^L$. چون نمادها دارای متوسط مدّت t هستند بایستی $T = Lt$ برقرار باشد. باجایگزینی $N(T)$ و T در معادلهٔ (۱.۴)، تعریف (۱.۴) متناظر با معادلهٔ (۲۴.۴) می‌شود. بنابراین، تعریف (۲.۴) تعمیم یافتهٔ تعریف (۱.۴) است.

ظرفیت کانال نوفه‌دار به وضوح کمتر از ظرفیت کانال بدون نوفه است. اگر مقدار اطلاع $H(X)$ ارائه شده باشد به قسمی که $H(X) < C$، در این صورت می‌توان این اطلاع را با یک خطای کوچک دلخواه ε منتقل کرد. در موردی که $H(X) > C$، ممکن است کدی ساخت به قسمی که کانال بتواند اطلاع را منتقل کند، ولی با حفظ عدم‌حتمیت $H(X|Y) = H(X) - C + \varepsilon$ و ε کوچک دلخواه است. این نکته را در یکی از بخشهای بعدی بیشتر بررسی خواهیم کرد.

محاسبهٔ ظرفیت در مورد کانالهای نوفه‌دار عموماً کار ساده‌ای نیست و باید به روشهای عددی یا تقریبهای تحلیلی متوسل شد. با وجود این، برای کانال متقارن دودویی مذکور قبلی محاسبهٔ ظرفیت کاملاً ساده است.

مثال ۴. ۳

کانال متقارن دودویی را همانند آنچه که در مثال (۴.۲) ارائه شد در نظر بگیرید. برای نرخ ارسال R به دست آوردیم که

$$\mathrm{R} = -\beta \log \beta - (۱-\beta)\log(۱-\beta) + p\log p + (۱-p)\log(۱-p)$$

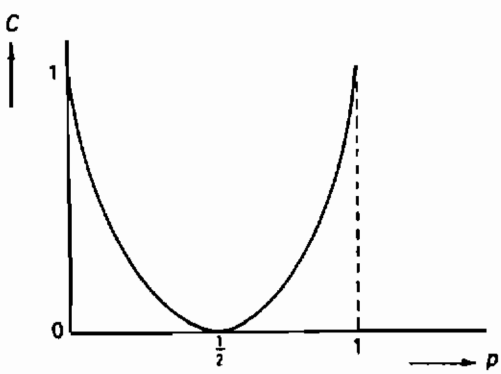
با $\beta = \alpha(۱-p) + (۱-\alpha)p$. بنابه تعریف ظرفیت، ظرفیت این کانال با انتخاب مقدار $p(x_۱) = \alpha_\circ$ که برای آن R ماکسیمم است به دست می‌آید. از این رو

$$C = \max_{\alpha} \mathrm{R},$$

که در آن R به صورت بالا داده شده است. چون p مستقل از α است، اگر $-\beta \log \beta - (۱-\beta)\log(۱-\beta)$ ماکسیمم شود R ماکسیمم خواهد شد. به سهولت دیده می‌شود که براساس ویژگیهای اندازه‌های اطلاع این عبارت به ازای $\beta = \beta_\circ = \frac{۱}{۲}$ ماکسیمم می‌شود. اکنون نتیجه می‌شود که $\alpha_\circ = \frac{۱}{۲}$، به قسمی که

$$C = \max_{\alpha} \mathrm{R} = ۱ + p\log p + (۱-p)\log(۱-p).$$

برای $p = \frac{۱}{۲}$ داریم $C = \circ$، به طوری که هیچ اطلاع منتقل نمی‌شود. از طرف دیگر برای $p = \circ$ یا $p = ۱$ داریم: $C = ۱$، یعنی $H(X)$ به طور کامل منتقل می‌شود. شکل (۸.۴) را ببینید. ▲



شکل ۸.۴- ظرفیت برای یک کانال متقارن دودویی

ظرفیت یک کانال دودویی را که لزومی ندارد متقارن باشد می‌توان به سادگی با یک روش نموداری تعیین کرد. ماتریس کانال چنین کانالی برابر است با

$$Q = \begin{bmatrix} q_{11} & q_{21} \\ q_{12} & q_{22} \end{bmatrix}.$$

فرض کنید $p(x_1) = \alpha$ و $q(y_1) = \beta$، و مطرح می کنیم

$$H_1 = -q_{11}\log q_{11} - q_{21}\log q_{21},$$

و

$$H_2 = -q_{12}\log q_{12} - q_{22}\log q_{22}.$$

در این صورت

$$H(Y) = -\beta\log\beta - (1-\beta)\log(1-\beta)$$

و

$$H(Y|X) = \alpha H_1 + (1-\alpha)H_2 = H_2 + \alpha(H_1 - H_2).$$

از این رو

$$\mathsf{R} = H(Y) - H_2 - \alpha(H_1 - H_2).$$

اکنون

$$\beta = \alpha q_{11} + (1-\alpha)q_{12} = q_{12} + \alpha(q_{11} - q_{12}),$$

یا

$$\alpha = \frac{\beta - q_{12}}{q_{11} - q_{12}}.$$

در شکل (۹.۴) طول DG مقدار $H(Y)$ را نمایش می‌دهد، و داریم: $AB = H_2$ و $KM = H_1$، از این رو $LM = H_1 - H_2$. اکنون نتیجه می‌شود

$$\frac{BE}{BL} = \frac{\beta - q_{12}}{q_{11} - q_{12}} = \alpha,$$

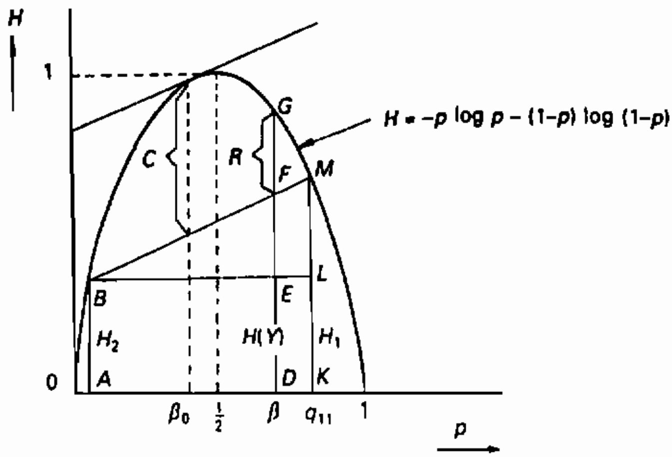
بنابراین

$$EF = \alpha(H_1 - H_2),$$

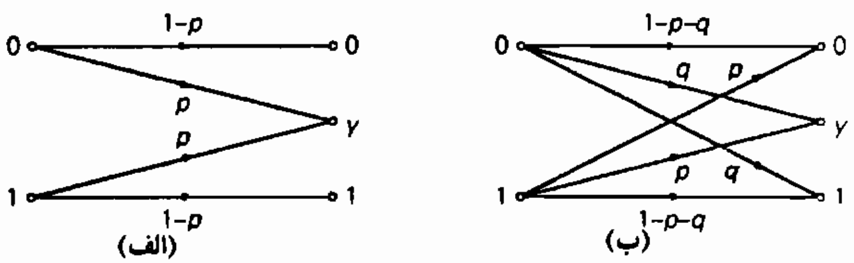
به قسمی که FG مقدار R را نشان می‌دهد. همچنین ظرفیت برابر ماکسیمم مقدار R است، R را با رسم خطّی موازی BM و مماس بر منحنی پیدا می کنیم. این نقطه تماس مقدار β_0 از β را که متعلّق به ظرفیت است تعیین می کند. بنابراین توزیع احتمال بهینهٔ منبع عبارت است از

$$p(x_1)=\frac{\beta_0-q_{12}}{q_{11}-q_{12}}=\alpha_0 .$$

معمولاً در کانال دودویی فرض می‌شود که $x_1=y_1=0$ و $x_2=y_2=1$. اگر الفبای خروجی سه‌سه‌ای باشد، برای مثال $Y=\{0,1,y\}$، که در آن y نمادی را نشان می‌دهد که اگر گیرنده قادر نباشد که ۰ یا ۱ را تشخیص دهد انتخاب می‌شود، در این صورت یک ***کانال پاک‌شدگی دودویی*** (BEC) پدید می‌آید. در این حالت هیچ تعویضی از نمادهای ۰ و ۱ نمی‌تواند رخ دهد (شکل (۴. ۱۰-الف) که یک BEC متقارن را نمایش می‌دهد ببینید). شکل (۴. ۱۰-ب) یک کانال پاک‌شدگی را که همچنین ممکن است موسوم به: ***کانال خطاها و پاک‌شدگی*** باشد نمایش می‌دهد.



شکل ۴.۹- ظرفیت برای یک کانال دودویی

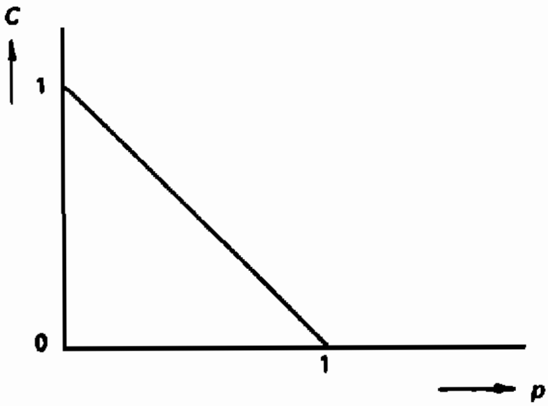


شکل ۴. ۱۰- (الف) کانال پاک‌شدگی دودویی متقارن؛ (ب) کانال خطاها و پاک‌شدگی

ماتریس کانال برای BEC متقارن عبارت است از

$$Q=\begin{bmatrix} 1-p & p & 0 \\ 0 & p & 1-p \end{bmatrix}.$$

ظرفیت این کانال برابر $p-1$ است (شکل (۱۱.۴) و همچنین تمرین (۷.۴) را ببینید).



شکل ۱۱.۴- ظرفیت کانال پاک‌شدگی دودویی متقارن

قضیهٔ ۲.۴

ظرفیت هر کانال نوفه‌دار متقارن گسسته به صورت زیر می‌باشد

$$C=\sum_{j=1}^{n} q(y_i|x_i)\log q(y_i|x_i)+\log n, \qquad (۲۵.۴)$$

و با استفاده از توزیع هم‌احتمال برای الفبای ورودی به دست می‌آید.

برهان

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

$$=\max_{p(x)}[\sum_{i=1}^{m} p(x_i)\sum_j q(y_i|x_i)\log q(y_i|x_i)-\sum_{j=1}^{n} q(y_i)\log q(y_i)]. \qquad (۲۶.۴)$$

به علّت تقارن $\sum_j q(y_i|x_i)\log q(y_i|x_i)$ برای تمام iها یکسان است. از این رو انتخاب احتمالهای ورودی تأثیری بر اوّلین جمله ندارد؛ بنابراین

$$C=\sum_{j=1}^{n} q(y_j|x_i)\log q(y_j|x_i)-\max_{p(x)}[\sum_{j=1}^{n} q(y_j)\log q(y_j)]. \qquad (۲۷.۴)$$

اگر احتمال ورودی طوری باشد که خروجی هم‌توزیع باشد جملهٔ دوم را می‌توان به

دست آورد. به واسطهٔ تقارن ماتریس انتقال Q، اگر احتمالهای ورودی هم‌شانس باشند این حالت را خواهیم داشت. ■

به عنوان آخرین مثال **کانال-** Z را بیان می کنیم. برای این الگو کانال دودویی فرض شده است که دقیقاً یکی از دو نماد ورودی مثلاً x_1 بدون خطا ارسال شده است؛ در حالی که نماد دیگر ممکن است با احتمال p به طور نادرست دریافت شود. الگوی کانال در شکل (۱۲.۴) نمایش داده شده است. ماتریس کانال عبارت است از

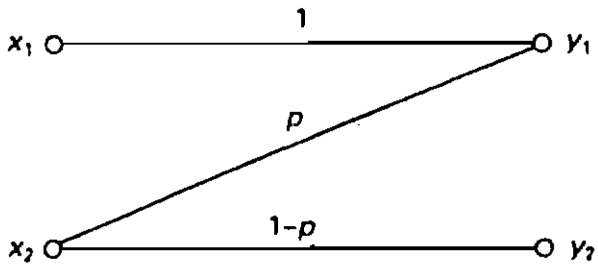
$$Q=\begin{bmatrix} 1 & 0 \\ p & 1-p \end{bmatrix}.$$

چون این کانال متقارن نیست از این پس $p(x_1)=q(y_1)$ و $p(x_2)=q(y_2)$ برقرار نخواهد بود. اکنون ظرفیت را با این فرض که $p(x_1)=1-\alpha_0$ و $p(x_2)=\alpha_0$ تعیین می کنیم، سپس نرخ ارسال R را محاسبه کرده و ماکسیمم R را به دست می‌آوریم. محاسبات در تمرین (۸.۴) داده شده‌اند. ظرفیت کانال- Z عبارت است از

$$C=\alpha_0[p\log p-(1-p)\log\alpha_0]-[1-\alpha_0(1-p)]\log[1-\alpha_0(1-p)], \qquad (۲۸.۴)$$

که در آن

$$\alpha_0=\frac{1}{1-p+p^{-p/(1-p)}}. \qquad (۲۹.۴)$$



شکل ۱۲.۴- کانال- Z

همان‌طور که از این آخرین الگوی کانال و تمرین (۸.۴) پیداست محاسبهٔ ظرفیت کانال برای کانالهای پیچیده‌تر به سرعت پیچیده می‌شود. در این صورت باید ظرفیت را به طریق عددی تعیین کرد.

۳.۴ احتمال خطا و ابهام

همان‌طور که در بخش (۲.۴) قبلاً یادآوری شد، ممکن است نمادها در حین انتقال در کانال ارتباطی به علّت اثر نوفه تغییر کنند که خطا حاصل می‌شود. علاوه بر ظرفیت کانال، کیفیت کانال ارتباطی نیز با این احتمال که خطایی در حین ارسال انجام شده است تعیین خواهد شد. می‌توان (متوسط) احتمال خطا را به دو روش، یعنی از دیدگاه گیرنده و فرستنده بیان کرد. از این پس فرض خواهیم کرد که تعداد نمادها در کانال ورودی برابر تعداد نمادها در خروجی است، بنابراین یک ماتریس کانال مربع داریم. اگر نماد y_j دریافت شده باشد، در این صورت اگر نماد x_i، $i \neq j$ ارسال شده باشد خطایی انجام شده است. این احتمال خطا برابر است با

$$p(e|y_j) = \sum_{\substack{i=1 \\ i\neq j}}^{n} p(x_i|y_j) = 1 - p(x_j|y_j). \qquad (۳۰.۴)$$

با میانگین گرفتن روی تمام نمادهای دریافت شده به دست می‌آوریم

$$P_e = \sum_{j=1}^{n} q(y_j)\cdot p(e|y_j) = \sum_{j=1}^{n} q(y_j)[1 - p(x_j|y_j)]. \qquad (۳۱.۴)$$

به طور مشابه، از نقطه‌نظر فرستنده می‌توان متوسط احتمال خطا را به صورت زیر به دست آورد

$$P_e = \sum_{i=1}^{n} p(x_i)[1 - q(y_i|x_i)]. \qquad (۳۲.۴)$$

این موضوع که متوسط احتمالهای خطا برابرند در قضیهٔ زیر نشان داده خواهد شد:

قضیهٔ ۳.۴

برای یک کانال ارتباطی با ماتریس کانال مربع متوسط احتمال خطا از دیدگاه گیرنده و از دیدگاه فرستنده با هم برابرند.

برهان

در واقع، نشان می‌دهیم که سمت راست معادلهٔ (۳۱.۴) و معادلهٔ (۳۲.۴) با یکدیگر برابرند. با توجّه به معادلهٔ (۳۱.۴) داریم

$$P_e = \sum_{j=1}^{n} p(y_j)[1 - p(x_j|y_j)] = \sum_{j=1}^{n} q(y_j)[1 - \frac{r(x_j, y_j)}{q(y_j)}]$$
$$= 1 - \sum_{j=1}^{n} r(x_j, y_j).$$

چون بنا به تعریف داریم

$$\sum_{i=1}^{n}\sum_{j=1}^{n} r(x_i, y_j) = 1,$$

نتیجه می‌شود که

$$P_e = \sum_{i=1}^{n}\sum_{j=1}^{n} r(x_i, y_j) - \sum_{j=1}^{n} r(x_j, y_j) = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} r(x_i, y_j)$$

$$= \sum_{i=1}^{n} p(x_i) \sum_{\substack{j=1 \\ j\neq i}}^{n} q(y_j | x_i).$$

و سرانجام داریم

$$P_e = \sum_{i=1}^{n} p(x_i)[1 - q(y_i | x_i)].$$

این همان فرمول (۳۲.۴) می‌باشد. ■

پس همان‌طور که در اثبات قضیهٔ (۳.۴) نشان داده شد، متوسط احتمال خطا را نیز می‌توان به صورت زیر نوشت

$$p_e = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} r(x_i, y_j). \tag{۳۳.۴}$$

رابطه‌ای بین ابهام $H(X|Y)$ و متوسط احتمال خطا P_e وجود دارد. این رابطه را نیز به نام ***نابرابری فانو*** می‌شناسند.

قضیهٔ ۴.۴ (نابرابری فانو)

فرض کنید $H(P_e)$ به صورت زیر تعریف شده باشد

$$H(P_e) = -P_e \log P_e - (1 - P_e)\log(1 - P_e),$$

که در آن P_e متوسط احتمال خطاست، همان‌طور که در معادلهٔ (۳۳.۴) داده شده است. اکنون نابرابری زیر برقرار است

$$H(X|Y) \leq H(P_e) + P_e \log(n-1). \tag{۳۴.۴}$$

برهان

سمت راست نابرابری فرمول (۳۴.۴) را می‌توان به کمک معادلهٔ (۳۳.۴) به صورت زیر نوشت

$$H(P_e)+P_e\log(n-1)=P_e\log\left(\frac{n-1}{P_e}\right)+(1-P_e)\log\left(\frac{1}{1-P_e}\right)$$
$$=\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} r(x_i,y_j)\log\left(\frac{n-1}{P_e}\right)+\sum_{i=1}^{n} r(x_i,y_i)\log\left(\frac{1}{1-P_e}\right). \qquad (۳۵.۴)$$

می‌توان ابهام را به همین طریق برحسب مجموعهای مشابهی نوشت

$$H(X|Y)=-\sum_{i=1}^{n}\sum_{j=1}^{n} r(x_i,y_j)\log p(x_i|y_j)$$
$$=\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} r(x_i,y_j)\log\left(\frac{1}{p(x_i|y_j)}\right)+\sum_{i=1}^{n} r(x_i,y_j)\log\left(\frac{1}{p(x_i|y_i)}\right). \qquad (۳۶.۴)$$

با کم کردن معادلهٔ (۳۵.۴) از معادلهٔ (۳۶.۴) نتیجه می‌شود که

$$H(X|Y)-H(P_e)-P_e\log(n-1)$$
$$=\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} r(x_i,y_j)\log\left\{\frac{P_e}{(n-1)p(x_i|y_j)}\right\}+\sum_{i=1}^{n} r(x_i,y_i)\log\left\{\frac{1-P_e}{p(x_i|y_i)}\right\}. \qquad (۳۷.۴)$$

با تبدیل به لگاریتم طبیعی و به کارگیری نابرابری

$$\ln a\le a-1$$

برای سمت راست معادلهٔ (۳۷.۴) نتیجه می‌شود که

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} r(x_i,y_j)\log\left\{\frac{P_e}{(n-1)p(x_i|y_j)}\right\}+\sum_{i=1}^{n} r(x_i,y_i)\log\left\{\frac{1-P_e}{p(x_i|y_i)}\right\}$$
$$\le\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} r(x_i,y_j)\frac{1}{\ln 2}\left\{\frac{P_e}{(n-1)p(x_i|y_j)}-1\right\}+\sum_{i=1}^{n} r(x_i,y_i)\frac{1}{\ln 2}\left\{\frac{1-P_e}{p(x_i|y_j)}-1\right\}$$
$$=\frac{1}{\ln 2}\left\{\left[\left\{\frac{P_e}{n-1}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{r(x_i,y_j)}{p(x_i|y_j)}\right\}-\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} r(x_i,y_j)+\sum_{i=1}^{n}\left\{(1-P_e)\sum_{i=1}^{n}\frac{r(x_i,y_i)}{p(x_i|y_i)}\right\}-\sum_{i=1}^{n} r(x_i,y_i)\right]\right\}$$
$$=\frac{1}{\ln 2}\left\{\frac{P_e}{n-1}.(n-1)-\left[1-\sum_{i=1}^{n} r(x_i,y_i)\right]+(1-P_e)-\sum_{i=1}^{n} r(x_i,y_i)\right\}=0.$$

از

$$H(X|Y) - H(P_e) - P_e \log(n-1) \le 0,$$

اکنون فرمول (۳۴.۴) مستقیماً به دست می‌آید. ■

برای این‌که بدانیم تحت چه شرایطی برابری در نابرابری فانو رخ می‌دهد، بـه خـاطر بیاورید که نابرابری $\ln a \le a-1$ به تساوی تبدیل می‌شود اگر و تنهـا اگـر $a=1$. در ایـن صورت بدون ارائهٔ اثبات شرط زیر باید برقرار باشد:

$$p(x_i|y_j) = \frac{P_e}{n-1} \quad \text{برای تمام } i \text{ و } j \text{ها به استثنای } i=j \qquad (۳۸.۴)$$

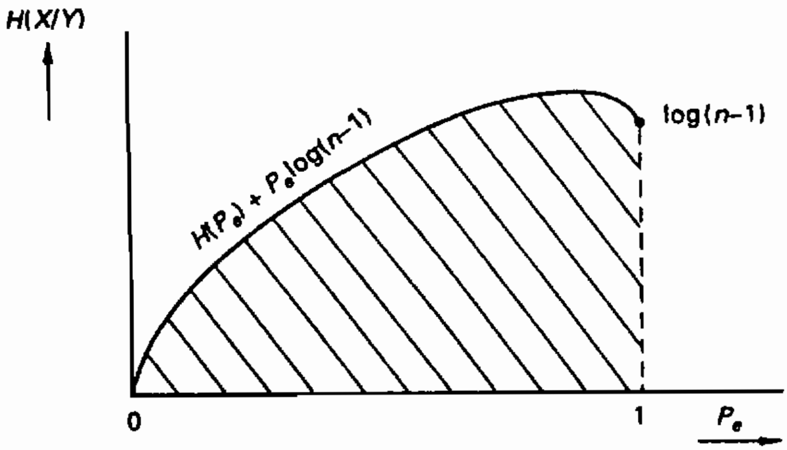
و

$$p(x_i|y_i) = 1 - P_e \quad \text{برای تمام } i \text{ها} \qquad (۳۹.۴)$$

چون برای تمام jها

$$\sum_{i=1}^{n} p(x_i|y_j) = 1,$$

می‌توان نتیجه گرفت که شرط مذکور در (۳۹.۴) قبلاً با معادلهٔ (۳۸.۴) سـاده شـده اسـت. معادله (۳۸.۴) ایجاب می‌کند که برای نماد خروجی داده شده تمـام نمادهـای ورودی بـه استثنای نماد انتخاب شده احتمال رخ‌داد مساوی دارند.



شکل ۱۳.۴- رابطهٔ بین احتمال خطا و ایهام

علاوه بر اثبات رسمی، نابرابری فانو را می‌توان به صورت زیر بررسی کرد. متوسـط

عدم‌حتمیت دربارهٔ X را اگر Y معلوم باشد می‌توان به عنوان عدم‌حتمیتی که خواه خطایی انجام شده یا نشده باشد در نظر گرفت، و اگر خطایی (با احتمال P_e) انجام شده باشد در این صورت عدم‌حتمیت $(n-۱)$ نماد باقی‌مانده ارسال شده است. اولین مقدار عدم‌حتمیت برابر با $H(P_e)$ است، در صورتی که دومین مقدار آن حداکثر برابر $\log(n-۱)$ است که با احتمال P_e وزن‌دار شده است. از این نابرابری نتیجه می‌شود که متوسط احتمال خطای کوچک به این معناست که ابهام کوچک است. نابرابری به صورت نمودار در شکل (۱۳.۴) رسم شده است.

۴.۴ قضیهٔ کدگذاری برای کانالهای بی‌حافظهٔ گسسته

در حالت کانال نوفه‌دار پیام ارسال شده می‌تواند با خطا همراه باشد. با وجود این، روشن است که با ارسال اطلاع به صورت حشو مثلاً تکرار پیام می‌توان احتمال خطا را کاهش داد. می‌توان انتظار داشت که برای این‌که احتمال خطا به صفر میل کند، حشو مورد نیاز آن قدر بزرگ است که نرخ ارسال نیز به صفر میل می‌کند. نشان ویژهٔ قضیهٔ زیر آن است که ممکن است اطلاع را با نرخ C با احتمال خطا یا ابهام به اندازهٔ دلخواه کوچک و از طریق کانال ارسال کرد.

این مطلب برای نرخ بزرگتر از C درست نیست. اگر اطلاع را با نرخ $C+R'$ ارسال کنیم در این صورت لزوماً ابهامی بزرگتر یا مساوی R' وجود خواهد داشت.

قضیهٔ ۵.۴ (قضیهٔ دوم کدگذاری شانون)

اگر $H(X) \le C$ آن‌گاه از طریق کانال بی‌حافظه‌ای با ظرفیت C ارسال مقدار اطلاع $H(X)$ با احتمال خطایی به اندازهٔ دلخواه کوچک (ابهام کوچک) امکان‌پذیر است.

اگر $H(X) > C$ امکان دارد منبع را به طریقی رمزگذاری کرد که ابهام کمتر از $H(X)-C+\varepsilon$ شود که در آن ε به اندازهٔ دلخواه کوچک است. هیچ روش کدگذاری وجود ندارد که ابهامی کمتر از $H(X)-C$ ایجاد کند.

برهان

قضیه ساخت روشهای کدگذاری را که دارای ویژگیهای مطلوب باشد نمی‌خواهد بلکه تنها دنبال وجود چنین کدهایی است.

اکنون منبعی با مقدار اطلاع $H(X)$ در نظر بگیرید. فرض کنید توزیع احتمال منبع به

قسمی است که ظرفیت C حاصل می‌گردد؛ یعنی $C = H(X) - H(X|Y)$. تعداد محتملترین پیامها، به طول l برابر است با $M_x = ۲^{lH(X)}$؛ همهٔ این پیامها احتمالی برابر دارند. احتمال کل بقیهٔ پیامها کوچک است. به طور مشابه، تعداد پیامهای دریافت شده با احتمال زیاد در طرف دیگر کانال متعلّق به مجموعه‌ای با $M_y = ۲^{lH(Y)}$ پیام می‌باشد.

هر پیام دریافت شده $\tilde{y}$ به علّت تغییر شکل در کانال نوفه‌دار می‌توانند از تعدادی پیامهای $\tilde{x}$ ورودی آمده باشد. محتملترین تعداد ورودیهایی که به یک پیام دریافت شده ختم می‌شوند برابر است با $M_{x|y} = ۲^{lH(X|Y)}$.

در حالت تنظیم مطلوب منبع با کانال تعداد پیامهای احتمالی ارسال شده برابر با $M_C = ۲^{lC}$ است. در حالت عدم تنظیم مطلوب $\mathsf{R} < C$، و از این رو $M_\mathsf{R} = ۲^{l\mathsf{R}} < M_C$. اکنون کانالی با $\mathsf{R} < C$ و همان منبع اطلاع را در نظر می‌گیریم.

گزیده‌ای از ورودیهای ممکن کانال متعلّق به مجموعهٔ M_x را M_R می‌نامیم و نشان می‌دهیم که امکان‌پذیر است که احتمال خطا را به قدر ممکن کوچک کنیم.

احتمال آن که یک پیام $\tilde{x}_i \in M_x$ ارسال شده باشد با احتمال این که این پیام $\tilde{x}_i$ به M_R متعلّق باشد برابر است، این احتمال برابر است با

$$P(\tilde{x}_i \in M_\mathsf{R}) = \frac{۲^{l\mathsf{R}}}{۲^{lH(X)}} = ۲^{l\{\mathsf{R} - H(X)\}},$$

که (تقریباً) برای همهٔ پیامهای $\tilde{x}_i \in M_x$ به یک اندازه بزرگ است. سپس یک پیام دریافت شدهٔ $\tilde{y}_i$ را در نظر می‌گیریم. تعداد پیامهای ارسال شده‌ای را که به طور متوسط می‌توانند به پیام دریافت شدهٔ یکسان $\tilde{y}_i$ منجر شوند با $M_{x|y}$ نشان می‌دهیم. اکنون به تصادف، پیام $\tilde{x}_i$ را انتخاب می‌کنیم. اگر علاوه بر $\tilde{x}_i$ یک یا بیشتر از یک کدواژه $\tilde{x}_j$ که متعلّق به هر دوی $M_{x|y}$ و M_R است وجود داشته باشد و بتوانند به نماد دریافت شده یکسان $\tilde{y}_i$ منجر شوند در این صورت می‌تواند خطایی رخ دهد. بنابراین احتمال یک خطا برابر است با

$$P_e = P\{\text{حداقل یک } \tilde{x}_j \text{، } j \neq i \text{ به } (M_{x|y} \cap M_\mathsf{R}) \text{ متعلّق باشد}\}. \qquad (۴۰.۴)$$

این احتمال اجتماعی از تعدادی پیشامدهاست و از نظریهٔ احتمال می‌دانیم که کوچکتر یا مساوی با مجموع احتمالهای پیشامدهای تکی است. از این مطلب نتیجه می‌شود که

$$P_e \le \sum_{\substack{j=۱ \\ j \neq i}}^{M_{x|y}} P[(\tilde{x}_j \in M_{x|y}) \cap (\tilde{x}_j \in M_\mathsf{R})]. \qquad (۴۱.۴)$$

احتمال این که پیامی متعلّق به $M_{x|y}$ بوده و ارسال نیز شده باشد برابر است با

$$P[(\tilde{x}_j \in M_{x|y}) \cap (\tilde{x}_j \in M_\mathrm{R})] = P(\tilde{x}_j \in M_{x|y}) \cdot P(\tilde{x}_j \in M_\mathrm{R})$$
$$\leq P(\tilde{x}_j \in M_\mathrm{R}), \qquad (۴۲.۴)$$

چون بنابر تعریف $P(\tilde{x}_j \in M_{x|y}) \leq ۱$، بنابراین

$$P_e \leq \sum_{\substack{j=۱ \\ j \neq i}}^{M_{x|y}} P(\tilde{x}_j \in M_\mathrm{R}). \qquad (۴۳.۴)$$

با جایگزینی مقدار به دست آمدهٔ قبل برای $P(\tilde{x}_j \in M_\mathrm{R})$ نتیجه می‌شود

$$P_e \leq \sum_{\substack{j=۱ \\ j \neq i}}^{M_{x|y}} ۲^{l\{\mathrm{R}-H(X)\}},$$

یعنی:

$$P_e \leq \{M_{x|y} - ۱\} ۲^{l\{\mathrm{R}-H(X)\}}. \qquad (۴۴.۴)$$

چون $\mathrm{R} < C$، می‌توان ثابت کرد که

$$\mathrm{R} = C - \theta = H(X) - H(X|Y) - \theta, \qquad (۴۵.۴)$$

که در آن θ یک ثابت مثبت است. از این نتیجه می‌شود که

$$P_e \leq \{۲^{lH(X|Y)} - ۱\} ۲^{l\{-H(X|Y)-\theta\}} \leq ۲^{-l\theta}.$$

یعنی احتمال خطا را می‌توان با افزایش مقدار l به دلخواه کوچک کرد، به شرط آن که $\mathrm{R} < C$.
علاوه بر این، نابرابری فانو برای ایهام $H(X|Y)$ برقرار است

$$H(X|Y) \leq H(P_e) + P_e \log(n-۱).$$

از این رو اگر P_e به صفر نزدیک شود، ایهام نیز به صفر نزدیک خواهد شد. در این زمینه‌ها می‌توان نتیجه‌گیری کرد که اگر $H(X) \leq C$ با احتمال خطای کوچک قابل اغماضی ارسال امکان‌پذیر است.

قسمت دوم قضیه را می‌توان به صورت زیر ثابت کرد. اگر $H(X) > C$ از باقی‌ماندهٔ اطلاع صرف‌نظر خواهد شد. این درگیرنده ایهام $H(X|Y) > ۰$ را موجب می‌گردد. با وجود این، این ایهام حدّاقل برابر $H(X) - C$ خواهد بود. فرض می‌کنیم خلاف آن درست باشد،

$H(X|Y) \leq H(X)-C$. در این صورت برای δ مثبتی داریم $H(X|Y) \leq H(X)-C-\delta$ و از این رو

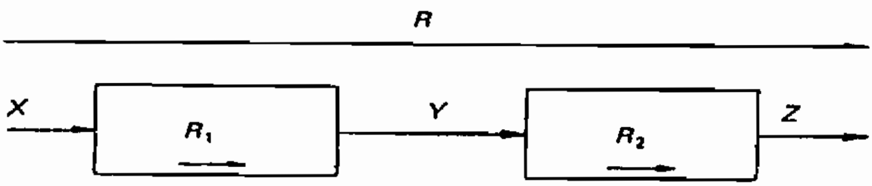
$$H(X)-H(X|Y)=C+\delta.$$

این با تعریف C به عنوان ماکسیمم $H(X)-H(X|Y)$ تناقض دارد. از این رو $H(X|Y) \geq H(X)-C$ و بنابراین برای مثال $H(X|Y)=H(X)-C+\varepsilon$ که در آن ε می‌تواند به دلخواه کوچک باشد. ■

در این جا توجّه کنید که با قضیهٔ (۵.۴) فقط ثابت کردیم که ارسال بدون خطا امکان‌پذیر است ولی مشخّص نشد که این کار را چگونه باید انجام داد. در عمل، کدهای تصحیح خطا برای تقلیل خطاهای کانال به کار برده می‌شوند. چون به دلایل عملی نمی‌توان طول پیامها و کدواژه‌های همراه آن را بی‌نهایت بزرگ اختیار کرد باید احتمال خطای معیّنی را پذیرفت. این مقدار با توجّه به کاربرد معمولاً از 10^{-3} تا 10^{-12} تغییر خواهد کرد.

۵.۴ کانالهای متوالی

در بسیاری از موارد، ارسال یا ذخیرهٔ اطلاع به طریقی رخ می‌دهد که الگوی استفاده شده تا کنون، که در آن تنها یک کانال داریم، خیلی ساده است. اکنون الگویی در نظر می‌گیریم که شامل دو یا بیشتر از دو کانال متوالی است. در این صورت اطلاع ارسال شده از طریق هر یک از کانالهای فرعی برابر اطلاع ورودی منهای ایهام مربوط به آن کانال فرعی می‌باشد. حال ممکن است از خود بپرسیم که چه رابطه‌ای بین نرخ کلّ ارسال و نرخ ارسال در هر کانال وجود دارد. فرض کنید که X ورودی کانال اوّل باشد در حالی که خروجی آن مجدّداً ورودی برای کانال دوم است. خروجی کانال دوم با Z داده می‌شود (شکل (۱۴.۴) را ببینید). مقدار اطلاع $H(X)$ در ورودی ارائه شده و $H(Z)$ در خروجی دریافت شده است، در حالی که $H(Y)$ بین دو بخش عبور می‌کند.



شکل ۱۴.۴- کانالهای متوالی

قضیهٔ ۶.۴

اگر یک کانال با نرخ ارسال R_1 توسط کانال دومی دنبال شود در این صورت برای نرخ کل ارسال R برای هر دو کانال داریم

$$\mathrm{R} \le \mathrm{R}_1. \tag{۴۶.۴}$$

برهان

باید ثابت کنیم که

$$\mathrm{R} = H(X) - H(X|Z) \le \mathrm{R}_1 = H(X) - H(X|Y).$$

یک نماد x_i از الفبای X یک نماد y_j از الفبای Y را به عنوان نتیجه خواهد داشت، که به نوبت خود نماد z_k را از الفبای Z به عنوان یک نتیجه خواهد داشت. از این رو z_k تنها از طریق y_j به x_i بستگی دارد، به قسمی که

$$p(z_k | y_j, x_i) = p(z_k | y_j),$$

برای تمام i، j و k. با استفاده از قضیهٔ بیز داریم

$$\frac{p(x_i | y_j, z_k) \cdot p(z_k | y_j)}{p(x_i | y_j)} = p(z_k | y_j),$$

یا

$$p(x | y, z) = p(x | y).$$

اکنون با استفاده از $\ln a \le a - 1$ یا $\log a \le (a-1)\log e$ به دست می‌آوریم

$$H(X|Z) - H(X|Y)$$

$$= -\sum_x \sum_z p(x,z) \log p(x|z) + \sum_x \sum_y p(x,y) \log p(x|y)$$

$$= -\sum_x \sum_y \sum_z p(x,y,z) \log p(x|z) + \sum_x \sum_y \sum_z p(x,y,z) \log p(x|y)$$

$$= -\sum_x \sum_y \sum_z p(x,y,z) \log \frac{p(x|z)}{p(x|y)} = -\sum_x \sum_y \sum_z p(x,y,z) \log \frac{p(x|z)}{p(x|y,z)}$$

$$= -\sum_y \sum_z p(y,z) \sum_x p(x|y,z) \log \frac{p(x|z)}{p(x|y,z)}$$

$$\geq -\sum_y \sum_z p(y,z)\log e \times \sum_x p(x|y,z)\left\{\frac{p(x|z)}{p(x|y,z)}-1\right\}$$

$$= -\sum_y \sum_z p(y,z)\log e \times \left\{\sum_x p(x|z) - \sum_x p(x|y,z)\right\}$$

$$= -\sum_y \sum_z p(y,z)\log e \times \{1-1\} = 0. \qquad (۴۷.۴)$$

از این رو ثابت شد که $H(X|Z) \geq H(X|Y)$، که از آن نتیجه می‌شود $R \leq R_1$. ■

قضیهٔ ثابت شده در بالا به‌نام **قضیهٔ پردازش داده‌ها** شناخته می‌شود. در واقع فرض نشان می‌دهد که وقتی داده‌ها متوالیاً پردازش شوند تنها اتلاف اطلاع امکان‌پذیر است. فقط اگر رابطهٔ یکتایی بین نمادهای ورودی و خروجی وجود داشته باشد می‌توان همهٔ اطلاع را حفظ کرد. در این صورت حالتی از یک کانال با ارسال بدون خطاست. با دقت بیشتری به قضیه، مشخّص می‌شود که کانالهای متوالی به ***اتلاف اطلاع منبع*** منجر می‌شود زیرا ابهام افزایش می‌یابد. امکان‌پذیر نیست که یک رابطهٔ ساده‌ای بین نرخهای ارسال R، R_1 و R_2 به‌دست آورد.

مثال ۴.۴

دو کانال متقارن دودویی متّصل‌شده به‌طور متوالی را در نظر بگیرید (شکل (۱۵.۴))، که در آن $0 \leq p \leq \frac{1}{2}$. در حالتی که برای هر دو نماد ورودی احتمال یکی است، برای کانال اوّل داریم

$$\begin{aligned} R_1 &= H(Y) - H(Y|X) \\ &= 1 - H(P), \end{aligned}$$

که در آن

$$H(P) = -p\log p - (1-p)\log(1-p).$$

نماد z_1 می‌تواند از x_1 به‌واسطهٔ y_1 یا y_2 به‌دست آید. بنابراین

$$p(z_1|x_1) = (1-p)^2 + p^2 = 1 - 2p(1-p) = 1 - p',$$

و

$$p(z_2|x_1) = 2p(1-p) = p'.$$

بدیهی است که این نیز درست است که $p(z_1) = p(z_2) = \frac{1}{2}$. در این صورت

$$\mathrm{R} = H(Z) - H(Z|X)$$
$$= ۱ - H(P')$$

که در آن

$$H(P') = -p' \log p' - (۱ - p') \log(۱ - p').$$

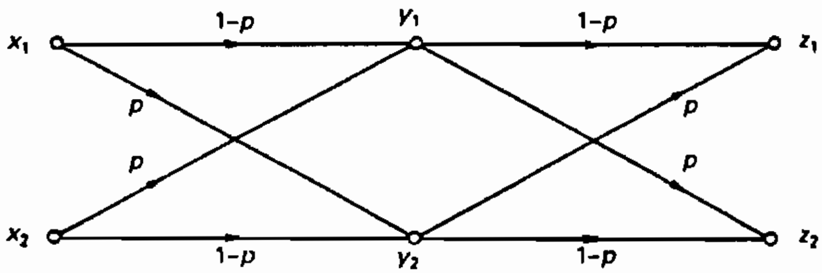
برای $۰ \le p \le \frac{۱}{۲}$ داریم $p \le p' = ۲p(۱-p) \le \frac{۱}{۲}$. بنابراین با ملاحظهٔ شکل (۱.۱) نتیجه‌گیری زیر امکان‌پذیر است

$$۰ \le H(P) \le H(P') \le ۱$$

و از این رو نتیجه می‌شود

$$\mathrm{R} \le \mathrm{R}_۱$$

هم‌زمان می‌توان مشاهده کرد که $H(Z|X) = H(Z|Y) + H(Y|X)$ نمی‌تواند برقرار باشد. زیرا $H(Z|Y) = H(Y|X) = H(P)$ و $H(Z|X) = H(P')$، این بدین معناست که $۲H(P) = H(P')$، که عموماً درست نمی‌باشد. از این رو نمی‌توان کلّ ایهام را از مجموع زیرایهامها به دست آورد، بلکه باید ماتریس کانال را برای همهٔ کانال تعیین کرد، از این رو در این حالت باید احتمالهای p' و $(۱-p')$ را تعیین کرد. ▲



شکل ۱۵.۴- دو کانال متقارن دودویی متوالی

۶.۴ کانالهای باحافظه

تا کنون کانالهای بی‌حافظه را بررسی کرده‌ایم. بدین معنی که فرض کردیم که رخ‌داد خطاها، یعنی تغییر نمادها در حین ارسال، مستقل از اشتباهات است. علاوه بر مزیّت تحلیل ساده این کانالها، به خصوص الگویی رضایت‌بخش می‌باشد. بنابراین اکثر تصحیح خطای کدها در عمل نیز براساس رخ‌داد خطاهای مستقل می‌باشند. هنوز در عمل تعداد روزافزونی

از حالتها با کانال باحافظه سروکار دارند. این موضوع مربوط به کاربرد نرخهای ارسال زیادتر است که به علّت نقایص در حین ارسال می‌تواند موجب دنباله‌ای از خطاهای متوالی گردد. این اثر با ذخیرهٔ داده‌ها در رسانه‌های نوری شماره‌ای و مغناطیسی نظیر دیسکها و نوارها نیز رخ می‌دهد زیرا وقوع آسیب یا عیب می‌تواند نسبتاً بیتهای بیشتر و بیشتری از طریق چگالی بیت بزرگتری بر واحد سطح تحریف کنند.

در حالتی که کانالها بی‌حافظه نیستند تعریف قوی‌تری از ظرفیت لازم است.

بلوکهای ورودی و خروجی با طول L را در نظر بگیرید. اکنون مانند ظرفیت برای نمادهای ورودی تکی داریم

$$C=\lim_{L\to\infty}\frac{1}{L}\max I(X_1,\ldots,X_L;Y_1,\ldots,Y_L) \tag{۴۸.۴}$$

که در آن روی همهٔ توزیعهای احتمال بلوکهای با طول L ماکسیمم گرفته می‌شود. به طور کلی ارزیابی کانالهای باحافظه خیلی مشکل است. در این جا خود را با کانالهای دودویی محدود می‌کنیم. وقوع دنباله‌ای از خطاهای وابسته را یک ***پراکندگی*** می‌نامیم. لزومی ندارد که دنباله‌ای از خطاهای متوالی رخ دهد؛ همچنین اغلب موقعیتهای موقتی را در نظر می‌گیریم که در آن احتمال خطای بزرگتر قابل توجّهی به عنوان یک پراکندگی وجود دارد. طول پراکندگی معمولاً حاکی از طول بین شروع خطای اوّل و انتهای آخرین خطا بدون توجّه به خطاهای میانی می‌باشد. برای بررسی رفتار این خطا می‌توان الگوهای توصیفی مبتنی بر پارامترهای آماری اندازه‌گیری شده از رخداد واقعی خطاها را به کار برد یا الگوهای مولدی برای کانالهای باحافظه ساخت که دنباله‌های خطایی که کم و بیش مشابه خطاهای کانال واقعی هستند تولید کند. در این جا چنین الگوی مولدی را بررسی خواهیم کرد که شامل زنجیر مارکوف با تعداد معیّنی حالت همراه با احتمالهای انتقال می‌باشد. ساده‌ترین الگو از این نوع الگوی ژیلبرت است که در شکل (۱۶.۴) رسم شده است.

الگوی ژیلبرت یک کانال باحافظه دارای دو حالت G و B (خوب و پراکنده) است و دنباله‌ای از صفرها و یکها را تولید می‌کند که در آن عدد ۱ یک خطا را نشان می‌دهد بنابراین صفر بی‌خطاست. یک بیت خطا همواره در حالت G (بدون خطا) صفر است، در حالی که در حالت B یک بیت خطا برابر صفر است با احتمال p و برابر یک است با احتمال $1-p$. زنجیر مارکوف هر زمان که به یک حالت جدید می‌رود یک خطا به وجود می‌آید. احتمالهای انتقال $P(B|G)$ و $P(G|B)$ آن قدر کوچکند که زنجیر مارکوف تمایل دارد که به ترتیب در حالت G یا B باقی بماند. همان‌طور که از الگو پیداست صفر

می‌تواند در هر حالت G و B تولید شده باشد. بنابراین برای تعیین احتمال خطا که در این جا با $P(۱)$ نشان داده می‌شود باید یک الگوی حالت، همان‌طور که در شکل (۱۷.۴) رسم شده است، معرّفی کنیم که در آن اکنون تنها حالت $B_۱$ متناظر با یک خطاست و $B_۰$ و G متناظر با رخ‌داد بدون خطا هستند.

می‌توان انتقال حالتهای G و B را به حالتهایی در شکل (۱۷.۴) به صورت $f(.)$ نشان داد به قسمی که

$$f(G)=۰ \qquad , \qquad f(B_۰)=۰ \qquad , \qquad f(B_۱)=۱.$$

برای زنجیر مارکوف شکل (۱۶.۴) برای احتمالهای دو حالت نتیجه می‌شود که

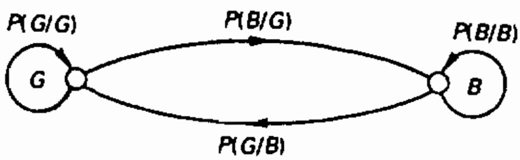
$$P(G)=\frac{P(G|B)}{P(B|G)+P(G|B)}$$

و

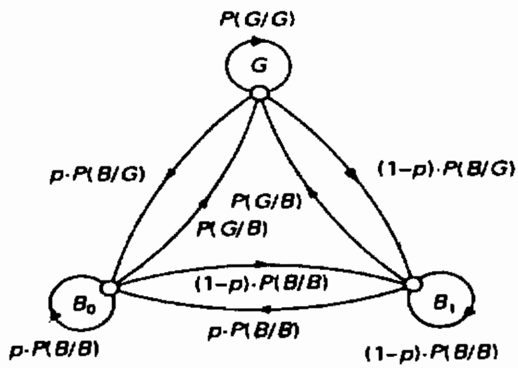
$$P(B)=\frac{P(B|G)}{P(B|G)+P(G|B)}.$$

خطایی با احتمال $(۱-p)$ در حالت B رخ می‌دهد بنابراین به دست می‌آوریم

$$P(۱)=(۱-p)P(B)=\frac{(۱-p)P(B|G)}{P(B|G)+P(G|B)}$$



شکل ۱۶.۴- الگوی ژیلبرت



شکل ۱۷.۴- الگوی ژیلبرت به صورت توسعه یافته

پس سه احتمال $P(1)$ رخ‌داد خطاها، $P(G|B)$، $P(B|G)$ و p الگوی ژیلبرت احتمال رخ‌داد خطاها را تعیین می کنند. گرچه الگوی ژیلبرت خودش ساده است، محاسبهٔ رفتار تفصیل‌شدهٔ دنباله‌های خطا نسبتاً پیچیده‌اند. چون زنجیر مارکوف مرتبهٔ اوّل تنها با دو حالت است کراندار می‌باشد، به خصوص نسبت به رفتار طول پراکندگی، دلیلی بر این که چرا گسترش حالتهای بیشتر مورد مطالعه قرار گرفته‌اند. علاوه بر این، بسط دیگری وجود دارد که در آن احتمال خطا در حالت G صفر نیست ولی مقدار مثبت (کوچکی) دارد.

۷.۴ تمرینها

۱.۴ دو نماد متفاوت x_1 و x_2 را در یک کانال ارتباطی با $p(x_1)=\frac{1}{2}$ می‌توان ارائه کرد. نمادهای y_1 و y_2 دریافت می‌شوند. در هر ثانیه یک نماد ارسال می‌شود. احتمالهای انتقال $q(y_j|x_i)=q_{ji}$ این کانال متقارن دودویی (BSC) با ماتریس انتقال زیر داده شده‌اند:

$$Q=\begin{bmatrix}\frac{2}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{2}{3}\end{bmatrix}.$$

(الف) عبارتی برای متوسط اطلاع متقابل $I(X;Y)$، که تنها به احتمال $p(x_i)$ نماد x_i و احتمال شرطی $q(y_j|x_i)$ نماد y_j به شرط x_i وابسته است، بیابید.

(ب) اطلاع متقابل $I(X;Y)$ را با کمک عبارت یافته‌شده محاسبه کنید.

(پ) مقدار اطلاع در طرف فرستنده را حساب کنید.

(ت) ابهام را محاسبه کنید.

(ث) اطلاع متقابل $I(X;Y)$ را با به کارگیری نتایج (پ) و (ت) محاسبه کنید.

۲.۴ هر ثانیه یکی از دو نماد x_i، $i=1,2$ به کانال ارتباطی با $p(x_1)=\alpha$ داده می‌شود. در طرف دریافت کننده انتخابی بین سه نماد y_j، $j=1,2,3$، می‌توان انجام داد که در آن احتمالهای انتقال $q(y_j|x_i)=q_{ji}$ توسط ماتریس انتقال زیر داده شده است:

$$Q=\begin{bmatrix}q_{11} & q_{21} & q_{31}\\ q_{12} & q_{22} & q_{32}\end{bmatrix}=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4}\end{bmatrix}.$$

(الف) مقدار اطلاع نمادها در طرف گیرنده را محاسبه کنید.
(ب) عدم‌حتمیت $H(Y|X)$ مربوط به نوفه را محاسبه کنید.
(پ) ظرفیت این کانال را محاسبه کنید.

۳.۴ سه نماد متفاوت $(x_۱, x_۲, x_۳)$ هر یک با احتمال $\frac{۱}{۳}$ را می‌توان با یک کانال ارتباطی ارسال کرد. سه نماد متفاوت $(y_۱, y_۲, y_۳)$ در طرف گیرنده مشاهده می‌شوند. احتمالهای انتقال $q(y_j|x_i) = q_{ij}$ توسط ماتریس انتقال زیر داده شده است:

$$Q = \begin{bmatrix} q_{۱۱} & q_{۲۱} & q_{۳۱} \\ q_{۱۲} & q_{۲۲} & q_{۳۲} \\ q_{۱۳} & q_{۲۳} & q_{۳۳} \end{bmatrix} = \begin{bmatrix} \frac{۱۰}{۱۶} & \frac{۲}{۱۶} & \frac{۴}{۱۶} \\ \frac{۵}{۱۶} & \frac{۶}{۱۶} & \frac{۵}{۱۶} \\ \frac{۶}{۱۶} & \frac{۱}{۱۶} & \frac{۹}{۱۶} \end{bmatrix}.$$

(الف) مقدار اطلاع دریافتی بر نماد چه قدر است؟
(ب) مقدار اطلاع انتقال یافته به سوی دیگر این کانال ارتباطی چه قدر است؟
(پ) احتمال ارسال $x_۲$ وقتی $y_۱$ دریافت شده باشد چه قدر است؟

۴.۴ یک منبع اطلاع سه نماد $x_۱$، $x_۲$ و $x_۳$ را تولید می‌کند. این منبع با یک کانال نوفه‌دار متصل است. دریافت کننده سه نماد $y_۱$، $y_۲$ و $y_۳$ را تشخیص می‌دهد. سیستم ارتباطی مذکور در بالا با ماتریس احتمالهای توأم $r(x_i, y_j) = r_{ij}$ با $i,j = ۱,۲,۳$ توصیف می‌شود:

$$\begin{bmatrix} r_{۱۱} & r_{۱۲} & r_{۱۳} \\ r_{۲۱} & r_{۲۲} & r_{۲۳} \\ r_{۳۱} & r_{۳۲} & r_{۳۳} \end{bmatrix} = \begin{bmatrix} \frac{۱}{۳۶} & \frac{۱}{۱۲} & ۰ \\ \frac{۱}{۴} & \frac{۱}{۹} & \frac{۵}{۳۶} \\ ۰ & \frac{۱}{۱۸} & \frac{۱}{۳} \end{bmatrix}.$$

با کمک داده‌های مفروض در بالا، مطلوب است:

(الف) مقدار اطلاع نمادهایی که در طرف دریافت کننده ارائه می‌شود؛
(ب) عدم‌حتمیت $H(Y|X)$ مربوط به نوفه؛
(پ) مقدار اطلاع انتقال داده شده به سوی دیگر این کانال نوفه‌دار.

۵.۴ یک کانال ارتباطی دودویی از دو زیرکانال متوالی ساخته شده است (شکل (۱۵.۴)

را ببینید).

احتمالهای انتقال $q(z_j|x_i)$، $i,j=۱,۲,\ldots$، از اوّلین زیرکانال توسط ماتریس زیر داده شده است

$$Q_{Z|X}=\begin{bmatrix}\frac{۳}{۴} & \frac{۱}{۴}\\ \frac{۱}{۴} & \frac{۳}{۴}\end{bmatrix}.$$

احتمالهای انتقال $q(y_j|z_i)$، $i,j=۱,۲,\ldots$، از دومین زیرکانال توسط ماتریس زیر داده شده است

$$Q_{Y|Z}=\begin{bmatrix}\frac{۲}{۳} & \frac{۱}{۳}\\ \frac{۱}{۳} & \frac{۲}{۳}\end{bmatrix}.$$

احتمالهای نمادهای x_i، $i=۱,۲,\ldots$ عبارتند از $\frac{۱}{۳}$ و $\frac{۲}{۳}$، و هر ثانیه یک نماد ارسال شده است.

(الف) $H(Z|X)$ را تعیین کنید.

(ب) مقدار اطلاع انتقال داده شده از X به Z را که با $R_۱$ نمایش داده می‌شود، تعیین کنید.

(پ) $H(Y|Z)$ را تعیین کنید.

(ت) مقدار اطلاع انتقال داده شده از X به Y را که با R نمایش داده می‌شود، تعیین کنید.

(ث) آیا انتظار دارید که R بزرگتر از $R_۱$ باشد یا کوچکتر؟ چرا؟

۶.۴ سه نماد متفاوت $(x_۱,x_۲,x_۳)$ هر یک با احتمال $\frac{۱}{۳}$ را می‌توان به یک کانال ارتباطی ارائه کرد. نمادهای $y_۱$، $y_۲$ و $y_۳$ دریافت می‌شوند. اگر نماد y_j دریافت شود وقتی نماد x_i، $i\neq j$ ارسال شده است در این صورت خطایی حاصل شده است. احتمالهای انتقال $q(y_j|x_i)=q_{ji}$ توسط ماتریس انتقال زیر داده شده است:

$$Q=\begin{bmatrix}۰/۵ & ۰/۳ & ۰/۲\\ ۰/۴ & ۰/۳ & ۰/۳\\ ۰/۱ & ۰/۹ & ۰\end{bmatrix}.$$

(الف) مقدار اطلاع $H(Y)$ دریافتی برای هر نماد در طرف گیرنده چه قدر است؟

(ب) اثر نوفه $H(Y|X)$ چه قدر است؟

(پ) احتمال خطای $p(e|y_۲)$ را در حالتی که نماد $y_۲$ دریافت شده است حساب کنید.

(ت) متوسط احتمال خطا را از نقطه‌نظر گیرنده محاسبه کنید.

(ث) متوسط احتمال خطا را از نقطه‌نظر فرستنده محاسبه کنید.

(ج) آیا می‌توانید مستقیماً از ماتریس انتقال ببینید که نتایج این کانال بد هستند.

(چ) مقدار اطلاع ارسال شده و ابهام را محاسبه کنید.

(ح) آیا اکنون نابرابری فانو برقرار است.

۷.۴ نمادهای یک منبع اطلاع دودویی با احتمال $p(x_۱)=\alpha$ و $p(x_۲)=۱-\alpha$ با کمک یک کانال پاک‌شدگی دودویی (BEC) با ماتریس انتقال زیر ارسال می‌گردد:

$$Q=\begin{bmatrix} ۱-p & p & ۰ \\ ۰ & p & ۱-p \end{bmatrix}.$$

(الف) عدم‌حتمیت $H(Y|X)$ مربوط به اثر نوفه را محاسبه کنید.

(ب) مقدار اطلاع ارسال شدهٔ R را تعیین کنید.

(پ) ظرفیت C این کانال را محاسبه کنید.

۸.۴ یک منبع دودویی با احتمالهای $p(x_۱)=(۱-\alpha)$ و $p(x_۲)=\alpha$ به یک کانال- Z متصل شده است (شکل (۱۲.۴) را ببینید).

(الف) مقدار اطلاع $H(Y)$ را بیابید.

(ب) مقدار اطلاع ارسال شده R را محاسبه کنید.

(پ) ظرفیت C این کانال را تعیین کنید.

۸.۴ جوابها

۱.۴ (الف) متوسط اطلاع متقابل $I(X;Y)$ همهٔ نمادهای ممکن x و y عبارتند از:

$$I(X;Y)=\sum_{i=j}^{۲}\sum_{j=۱}^{۲} r(x_i,y_j)\log\frac{r(x_i,y_j)}{p(x_i)q(y_j)}$$

$$=\sum_{i=j}^{۲}\sum_{j=۱}^{۲} p(x_i)q(y_j|x_i)\log\frac{q(y_j|x_i)}{q(y_i)}$$

$$= \sum_{i=j}^{۲} \sum_{j=۱}^{۲} p(x_i) q(y_j | x_i) \log \frac{q(y_j | x_i)}{\sum_{k=۱}^{۲} p(x_k) q(y_j | x_k)}.$$

(ب) با جانشین کردن مقادیر داده شده در عبارت بالا نتیجه می‌شود

$$I(X;Y) = \frac{۱}{۲}\left[\frac{۲}{۳}\log\frac{\frac{۲}{۳}}{\frac{۱}{۲}\times\frac{۲}{۳}+\frac{۱}{۲}\times\frac{۱}{۳}} + \frac{۱}{۳}\log\frac{\frac{۱}{۳}}{\frac{۱}{۲}\times\frac{۱}{۳}+\frac{۱}{۲}\times\frac{۲}{۳}}\right]$$

$$+\frac{۱}{۲}\left[\frac{۱}{۳}\log\frac{\frac{۱}{۳}}{\frac{۱}{۲}\times\frac{۲}{۳}+\frac{۱}{۲}\times\frac{۱}{۳}} + \frac{۲}{۳}\log\frac{\frac{۲}{۳}}{\frac{۱}{۲}\times\frac{۱}{۳}+\frac{۱}{۲}\times\frac{۲}{۳}}\right]$$

$$= \frac{۲}{۳}\log\frac{۴}{۳} + \frac{۱}{۳}\log\frac{۲}{۳} = ۰٫۰۸ \text{ بیت / ثانیه}.$$

(پ) مقدار اطلاع $H(X)$ در طرف فرستنده برابر است با

$$H(X) = -\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۲}\log\frac{۱}{۲} = ۱ \text{ بیت / ثانیه}.$$

(ت) برای ابهام $H(X|Y)$ بنابر تعریف داریم

$$H(X|Y) = -\sum_{i=۱}^{۲}\sum_{j=۱}^{۲} q(y_i) p(x_i | y_j) \log p(x_i | y_j).$$

احتمال نماد y_j از عبارت زیر به دست می‌آید

$$q(y_i) = \sum_{i=۱}^{۲} p(x_i) q(y_j | x_i) = \sum_{i=۱}^{۲} p(x_i) q_{ji},$$

بنابراین

$$q(y_۱) = \frac{۱}{۲}\times\frac{۲}{۳} + \frac{۱}{۲}\times\frac{۱}{۳} = \frac{۱}{۲},$$

و

$$q(y_۲) = \frac{۱}{۲}\times\frac{۱}{۳} + \frac{۱}{۲}\times\frac{۲}{۳} = \frac{۱}{۲},$$

در حالی که احتمالهای پسین x_i با استفاده از فرمول بیز به دست می‌آیند:

$$p(x_۱ | y_۱) = \frac{p(x_۱) q_{۱۱}}{q(y_۱)} = \frac{\frac{۱}{۲}\times\frac{۲}{۳}}{\frac{۱}{۲}} = \frac{۲}{۳},$$

$$p(x_۲|y_۱)=\frac{۱}{۳} \qquad , \qquad p(x_۱|y_۲)=\frac{۱}{۳} \qquad , \qquad p(x_۲|y_۲)=\frac{۲}{۳}.$$

با جانشین کردن این مقادیر در معادلهٔ ابهام نتیجه می‌شود

$$H(X|Y)=-\frac{۱}{۲}\left[\frac{۲}{۳}\log\frac{۲}{۳}+\frac{۱}{۳}\log\frac{۱}{۳}\right]-\frac{۱}{۲}\left[\frac{۱}{۳}\log\frac{۱}{۳}+\frac{۲}{۳}\log\frac{۲}{۳}\right]$$

$$=-\frac{۲}{۳}\log\frac{۲}{۳}-\frac{۱}{۳}\log\frac{۱}{۳}=\text{۰٫۹۱ بیت / ثانیه}.$$

(ث) رابطهٔ زیر برای اطلاع متقابل برقرار است:

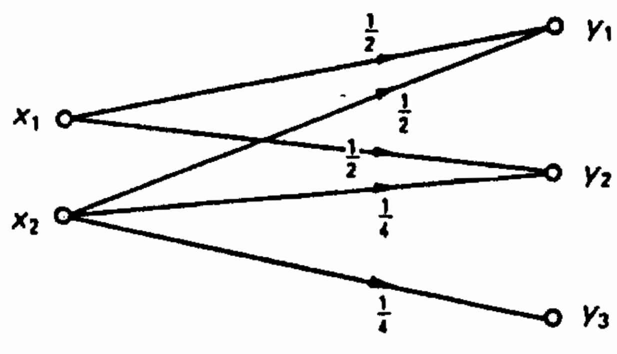
$$I(X;Y)=H(X)-H(X|Y).$$

با جای‌گزینی جوابهای حاصل در (پ) و (ت) در این عبارت داریم

$$I(X;Y)=۱-۰٫۹۱=\text{۰٫۰۹ بیت / ثانیه},$$

که با جواب به دست آمده در (ب) مطابق است.

۲.۴ (الف) کانال را می‌توان به صورت نشان داده شده در شکل (۱۸.۴) نمایش داد.



شکل ۱۸.۴- نمایش کانال تمرین (۲.۴)

$p(x_۱)=\alpha$ داده شده است لذا داریم $p(x_۲)=۱-\alpha$. اکنون احتمال رخ‌داد یک نماد y_j را می‌توان به صورت زیر تعیین کرد

$$q(y_j)=\sum_{i=۱}^{۲}p(x_i)q(y_j|x_i)=\sum_{i=۱}^{۲}p(x_i)q_{ji}.$$

از این نتیجه می‌شود

$$q(y_۱)=\frac{۱}{۲}\alpha+\frac{۱}{۲}(۱-\alpha)=\frac{۱}{۲},$$

$$q(y_2)=\frac{1}{2}\alpha+\frac{1}{4}(1-\alpha)=\frac{1+\alpha}{4},$$

$$q(y_3)=0\alpha+\frac{1}{4}(1-\alpha)=\frac{1-\alpha}{4}.$$

بنابراین، برای مقدار اطلاع دریافت شده $H(Y)$ داریم

$$H(Y)=-\frac{1}{2}\log\frac{1}{2}-\frac{1+\alpha}{4}\log\frac{1+\alpha}{4}-\frac{1-\alpha}{4}\log\frac{1-\alpha}{4}$$

$$=\frac{3}{2}-\frac{1+\alpha}{4}\log(1+\alpha)-\frac{1-\alpha}{4}\log(1-\alpha)$$ ثانیه / بیت.

(ب) بنا به تعریف برای $H(Y|X)$ داریم

$$H(Y|X)=-\sum_{i=1}^{2}\sum_{j=1}^{3}p(x_i)q(y_j|x_i)\log q(y_j|x_i)$$

$$=-\sum_{i=1}^{2}\sum_{j=1}^{3}p(x_i)q_{ji}\log q_{ji}$$

$$=-\alpha(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2})-(1-\alpha)(\frac{1}{2}\log\frac{1}{2}+\frac{1}{4}\log\frac{1}{4}+\frac{1}{4}\log\frac{1}{4})$$

$$=\alpha+(1-\alpha)\frac{3}{2}=\frac{3}{2}-\frac{\alpha}{2}$$ ثانیه / بیت.

(پ) مقدار اطلاع ارسال شده از رابطهٔ زیر به دست می‌آید

$$\mathrm{R}=H(Y)-H(Y|X)=\frac{3}{2}-\frac{1-\alpha}{4}\log(1+\alpha)-\frac{1-\alpha}{4}\log(1-\alpha)-\frac{3}{2}+\frac{\alpha}{2}$$

$$=-\frac{1+\alpha}{4}\log(1+\alpha)-\frac{1-\alpha}{4}\log(1-\alpha)+\frac{\alpha}{2}$$ بیت / ثانیه.

ظرفیت با انتخاب مقداری از $p(x_1)=\alpha$ که بازای آن R ماکسیمم می‌شود به دست می‌آید، یعنی

$$C=\max_{\alpha}\mathrm{R}.$$

این مقدار α از حلّ معادلهٔ $\frac{d\mathrm{R}}{d\alpha}=0$ به دست می‌آید:

$$\frac{dR}{d\alpha}=-\frac{1}{4}\log(1+\alpha)-\frac{\log e}{4}+\frac{1}{4}\log(1-\alpha)+\frac{\log e}{4}+\frac{1}{2}=0.$$

بنابراین داریم

$$\log\frac{1+\alpha}{1-\alpha}=2.$$

از این رو

$$\alpha = \frac{۳}{۵}.$$

با جانشین کردن در R حاصل می‌شود:

$$C = ۰٫۱۶ \text{ بیت / ثانیه}.$$

۳.۴ (الف) کانال را می‌توان به صورت شکل (۱۹.۴) نمایش داد. داریم

$$p(x_۱) = p(x_۲) = p(x_۳) = \frac{۱}{۳}$$

بنابراین احتمالهای $q(y_j)$، $j = ۱,۲,۳$ را به صورت زیر می‌توان محاسبه کرد:

$$q(y_j) = \sum_{i=۱}^{۳} p(x_i) q(y_j | x_i) = \sum_{i=۱}^{۳} p(x_i) q_{ji}.$$

از این نتیجه می‌شود

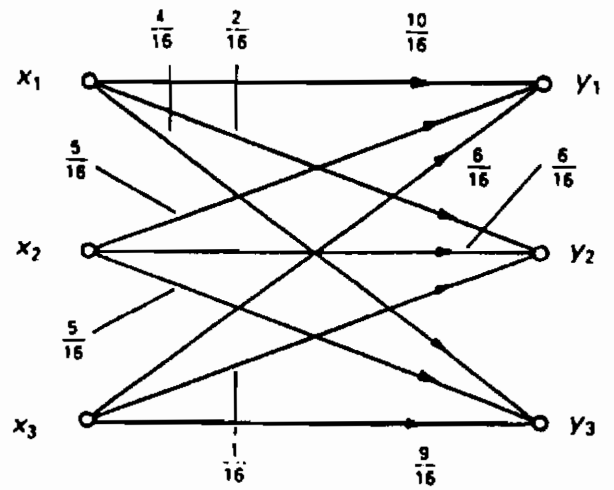
$$q(y_۱) = \frac{۱}{۳}\left(\frac{۱۰}{۱۶} + \frac{۵}{۱۶} + \frac{۶}{۱۶}\right) = \frac{۷}{۱۶}.$$

$$q(y_۲) = \frac{۱}{۳}\left(\frac{۲}{۱۶} + \frac{۶}{۱۶} + \frac{۱}{۱۶}\right) = \frac{۳}{۱۶}.$$

$$q(y_۳) = \frac{۱}{۳}\left(\frac{۴}{۱۶} + \frac{۵}{۱۶} + \frac{۹}{۱۶}\right) = \frac{۶}{۱۶}.$$

بنابراین، برای مقدار اطلاع دریافت شده داریم

$$H(Y) = -\frac{۷}{۱۶}\log\frac{۷}{۱۶} - \frac{۳}{۱۶}\log\frac{۳}{۱۶} - \frac{۶}{۱۶}\log\frac{۶}{۱۶} = ۱٫۵۱ \text{ بیت / نماد}.$$



شکل ۱۹.۴- نمایش کانال تمرین (۳.۴)

(ب) محاسبهٔ مقدار اطلاع ارسال شده بستگی به تعیین اثر نوفه $H(Y|X)$ دارد. این مقدار برابر است با

$$H(Y|X)=-\sum_{i=1}^{3}\sum_{j=1}^{3}p(x_i)q(y_j|x_i)\log q(y_j|x_i)=-\sum_{i=1}^{3}\sum_{j=1}^{3}p(x_i)q_{ji}\log q_{ji}$$

$$=-\frac{1}{3}\left(\frac{10}{16}\log\frac{10}{16}+\frac{2}{16}\log\frac{2}{16}+\frac{4}{16}\log\frac{4}{16}\right)-\frac{1}{3}\left(\frac{5}{16}\log\frac{5}{16}+\frac{6}{16}\log\frac{6}{16}+\frac{5}{16}\log\frac{5}{16}\right)$$

$$-\frac{1}{3}\left(\frac{6}{16}\log\frac{6}{16}+\frac{1}{16}\log\frac{1}{16}+\frac{9}{16}\log\frac{9}{16}\right)$$

$$=-\frac{1}{3}\times\frac{1}{16}(10\log 10+2+8+10\log 5+12\log 6+9\log 9-48\log 16).$$

با محاسبهٔ این عبارت نتیجه می‌شود

$$H(Y|X)=1{,}38 \text{ نماد / بیت}.$$

از این رو مقدار اطلاع ارسال شده برابر است با

$$\mathrm{R}=H(Y)-H(Y|X)=1{,}51-1{,}38=0{,}13 \text{ نماد / بیت}.$$

(پ) احتمال این که x_2 ارسال شده باشد وقتی y_1 دریافت شده است برابر احتمال پسین x_2 است. احتمال $q(y_1|x_2)$ را می‌دانیم. در این صورت با بهره‌وری از فرمول بیز نتیجه می‌شود که

$$p(x_2|y_1)=\frac{p(x_2)\cdot q(y_1|x_2)}{q(y_1)}=\frac{\frac{1}{3}\times\frac{5}{16}}{\frac{7}{16}}=\frac{5}{21}.$$

۴.۴ (الف) برای تعیین $H(Y)$، ابتدا باید به کمک ماتریس، $q(y_j)$، $j=1,2,3$، را محاسبه کرد. این احتمالها با جمع‌بندی احتمالهای توأم روی تمام iها به دست می‌آید:

$$\sum_{i=1}^{3}r(r_{ij})=q(y_j).$$

بنابراین داریم

$$q(y_1)=\frac{1}{36}+\frac{1}{4}=\frac{5}{18},$$

$$q(y_2)=\frac{1}{12}+\frac{1}{9}+\frac{1}{18}=\frac{1}{4},$$

$$q(y_3)=\frac{5}{36}+\frac{1}{3}=\frac{17}{36}.$$

اکنون

$$H(Y)=-\frac{۵}{۱۸}\log\frac{۵}{۱۸}-\frac{۱}{۴}\log\frac{۱}{۴}-\frac{۱۷}{۳۶}\log\frac{۱۷}{۳۶}=۱{,}۵۲ \text{ بیت / نماد.}$$

(ب) عدم‌حتمیت حاصل از نوفه عبارت است از $H(Y|X)$. دو روش برای تعیین $H(Y|X)$ وجود دارد. روش اوّل این است که $H(X,Y)$ را از ماتریس داده شده محاسبه کنیم و $H(X)$ را از آن کم کنیم؛ روش دوم این است که $q(y_j|x_i)$ را برای $i,j=۱,۲,۳$ محاسبه کنیم و سپس $H(Y|X)$ را مستقیماً از آن به دست آوریم. در هر دو حالت احتمالهای $p(x_i)$ باید تعیین شوند. روش اوّل محاسبات کمتری دارد. $p(x_i)$ با جمع‌بندی روی تمام jها در ماتریس داده شده به دست می‌آید

$$p(x_۱)=\frac{۱}{۳۶}+\frac{۱}{۱۲}=\frac{۱}{۹},$$
$$p(x_۲)=\frac{۱}{۴}+\frac{۱}{۹}+\frac{۵}{۳۶}=\frac{۱}{۲}, .$$
$$p(x_۳)=\frac{۱}{۱۸}+\frac{۱}{۳}=\frac{۷}{۱۸},$$

به طوری که

$$H(X)=-\frac{۱}{۹}\log\frac{۱}{۹}-\frac{۱}{۲}\log\frac{۱}{۲}-\frac{۷}{۱۸}\log\frac{۷}{۱۸}=۱{,}۳۸ \text{ بیت / نماد.}$$

به علاوه،

$$H(X,Y)=-\frac{۱}{۳۶}\log\frac{۱}{۳۶}-\frac{۱}{۱۲}\log\frac{۱}{۱۲}-\frac{۱}{۴}\log\frac{۱}{۴}-\frac{۱}{۹}\log\frac{۱}{۹}$$
$$-\frac{۵}{۳۶}\log\frac{۵}{۳۶}-\frac{۱}{۱۸}\log\frac{۱}{۱۸}-\frac{۱}{۳}\log\frac{۱}{۳}=۲{,}۴۵ \text{ بیت / نماد,}$$

بنابراین داریم

$$H(Y|X)=H(X,Y)-H(X)=۲{,}۴۵-۱{,}۳۸=۱{,}۰۷ \text{ بیت / نماد.}$$

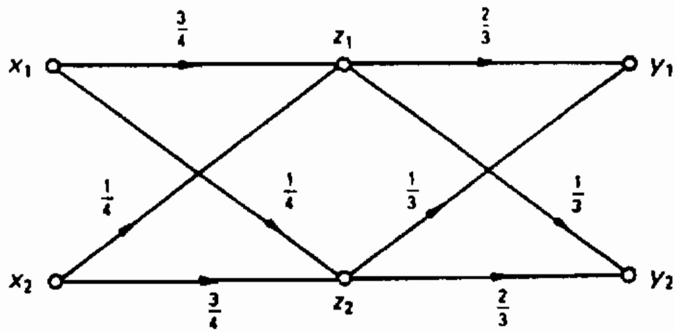
(پ) مقدار اطلاع ارسال شده R را می‌توان از رابطهٔ زیر به دست آورد.

$$\mathrm{R}=H(Y)-H(Y|X)=۱{,}۵۲-۱{,}۰۷=۰{,}۴۵ \text{ بیت / نماد.}$$

۵.۴ (الف) کانال را می‌توان به صورت شکل (۴. ۲۰) نمایش داد.

عدم‌حتمیت مربوط به نوفه $H(Z|X)$ با عبارت زیر محاسبه می‌شود

$$H(Z|X)=-\sum_{i=۱}^{۲}\sum_{j=۱}^{۲}p(x_i)q(z_j|x_i)\log q(z_j|x_i)$$



شکل ۴. ۲۰- کانالهای متوالی تمرین (۵.۴)

$$=-\frac{۱}{۳}\left[\frac{۳}{۴}\log\frac{۳}{۴}+\frac{۱}{۴}\log\frac{۱}{۴}\right]-\frac{۲}{۳}\left[\frac{۱}{۴}\log\frac{۱}{۴}+\frac{۳}{۴}\log\frac{۳}{۴}\right]$$

$$=-\frac{۳}{۴}\log\frac{۳}{۴}-\frac{۱}{۴}\log\frac{۱}{۴}=۰٫۸۱$$ بیت / نماد.

(ب) برای $R_۱$ داریم

$$R_۱=H(Z)-H(Z|X).$$

بنابراین ابتدا باید $H(Z)$ را محاسبه کنیم. داریم

$$q(z_۱)=p(x_۱)q(z_۱|x_۱)+p(x_۲)q(z_۱|x_۲)=\frac{۱}{۳}\times\frac{۳}{۴}+\frac{۲}{۳}\times\frac{۱}{۴}=\frac{۵}{۱۲},$$

و

$$q(z_۲)=p(x_۱)q(z_۲|x_۱)+p(x_۲)q(z_۲|x_۲)=\frac{۱}{۳}\times\frac{۱}{۴}+\frac{۲}{۳}\times\frac{۳}{۴}=\frac{۷}{۱۲},$$

بنابراین

$$H(Z)=-\frac{۵}{۱۲}\log\frac{۵}{۱۲}-\frac{۷}{۱۲}\log\frac{۷}{۱۲}=۰٫۹۷$$ بیت / نماد.

در نهایت از این نتیجه می‌شود

$$R_۱=۰٫۹۷-۰٫۸۱=۰٫۱۶$$ بیت / نماد.

(پ) برای $H(Y|Z)$ داریم

$$H(Y|Z)=-\sum_{i=۱}^{۲}\sum_{j=۱}^{۲}p(z_۱)q(y_j|z_i)\log p(y_j|z_i)$$

$$=-\frac{۵}{۱۲}\left[\frac{۲}{۳}\log\frac{۲}{۳}+\frac{۱}{۳}\log\frac{۱}{۳}\right]-\frac{۷}{۱۲}\left[\frac{۲}{۳}\log\frac{۲}{۳}+\frac{۱}{۳}\log\frac{۱}{۳}\right]$$

$$=-\frac{۲}{۳}\log\frac{۲}{۳}-\frac{۱}{۳}\log\frac{۱}{۳}=۰٫۹۲\ \text{بیت / نماد}.$$

(ت) برای احتمالهای خروجی کانال نتیجه می‌شود که

$$q(y_1)=q(z_1)q(y_1|z_1)+q(z_2)q(y_1|z_2)=\frac{۵}{۱۲}\times\frac{۲}{۳}+\frac{۷}{۱۲}\times\frac{۱}{۳}=\frac{۱۷}{۳۶},$$
$$q(y_2)=q(z_1)q(y_2|z_1)+q(z_2)q(y_2|z_2)=\frac{۵}{۱۲}\times\frac{۱}{۳}+\frac{۷}{۱۲}\times\frac{۲}{۳}=\frac{۱۹}{۳۶},$$

بنابراین داریم

$$H(Y)=-\frac{۱۷}{۳۶}\log\frac{۱۷}{۳۶}-\frac{۱۹}{۳۶}\log\frac{۱۹}{۳۶}=۰٫۹۹\ \text{بیت / نماد}.$$

برای تعیین $H(Y|X)$ می‌توان تصوّر کرد که دو کانال به طور سری با کانال جدیدی با نماد ورودی x و نماد خروجی y جای‌گزین شده باشد. در این صورت با تعیین مسیرهایی که می‌توان از یک نماد ورودی به همان نماد خروجی آمد و جمع کردن احتمالهای نظیرشان با یکدیگر، می‌توان ماتریس انتقال را به دست آورد. از این رو مثلاً داریم

$$q(y_1|x_1)=q(z_1|x_1)q(y_1|z_1)+q(z_2|x_1)q(y_1|z_2)=\frac{۳}{۴}\times\frac{۲}{۳}+\frac{۱}{۴}\times\frac{۱}{۳}=\frac{۷}{۱۲}.$$

با این روش برای چهار حالت ماتریس زیر حاصل می‌گردد:

$$\begin{bmatrix}\frac{۷}{۱۲} & \frac{۵}{۱۲}\\ \frac{۵}{۱۲} & \frac{۷}{۱۲}\end{bmatrix}.$$

در واقع فقط ماتریسها را در هم ضرب کرده‌ایم. اکنون اثر نوفه عبارت است از

$$H(Y|X)=-\frac{۷}{۱۲}\log\frac{۷}{۱۲}-\frac{۵}{۱۲}\log\frac{۵}{۱۲}=۰٫۹۷\ \text{بیت / نماد}.$$

بنابراین نتیجه می‌شود که

$$\mathrm{R}=H(Y)-H(Y|X)=۰٫۹۹-۰٫۹۷=۰٫۰۲\ \text{بیت / ثانیه}.$$

(ث) خواه یک نماد ورودی واقعاً منجر به یک خروجی متناظر بشود یا نشود عدم‌حتمیت برای دو کانال متوالی افزایش می‌یابد، چون اکنون تغییرات بیشتری امکان‌پذیر است. این کار، تعیین این را که کدام نماد ورودی موجب نماد خروجی مشاهده شده می‌باشد مشکلتر می‌کند. بنابراین

$$R\le R_1.$$

۶.۴ (الف) مقدار اطلاع $H(Y)$ را از احتمالهای $q(y_1)$، $q(y_2)$ و $q(y_3)$ می‌توان به دست آورد. داریم

$$q(y_j)=\sum_{i=1}^{3}p(x_i)q(y_j|x_i)=\sum_{i=1}^{3}p(x_i)q_{ji}.$$

از این رو

$$q(y_1)=\frac{1}{3}\left(\frac{5}{10}+\frac{4}{10}+\frac{1}{10}\right)=\frac{1}{3},$$

$$q(y_2)=\frac{1}{3}\left(\frac{3}{10}+\frac{3}{10}+\frac{9}{10}\right)=\frac{1}{2},$$

$$q(y_3)=\frac{1}{3}\left(\frac{2}{10}+\frac{3}{10}+0\right)=\frac{1}{6}.$$

بنابراین

$$H(Y)=-\frac{1}{3}\log\frac{1}{3}-\frac{1}{2}\log\frac{1}{2}-\frac{1}{6}\log\frac{1}{6}=1{,}46\ \text{بیت / نماد}.$$

(ب) برای اثر نوفه داریم

$$H(Y|X)=-\sum_{i=1}^{3}\sum_{j=1}^{3}p(x_i)q(y_j|x_i)\log q(y_j|x_i)$$

$$=-\frac{1}{3}[0{,}5\log 0{,}5+0{,}3\log 0{,}3+0{,}2\log 0{,}2]$$

$$-\frac{1}{3}[0{,}4\log 0{,}4+0{,}3\log 0{,}3+0{,}3\log 0{,}3]$$

$$-\frac{1}{3}[0{,}1\log 0{,}1+0{,}9\log 0{,}9]$$

$$=0{,}495+0{,}523+0{,}158=1{,}18\ \text{بیت / نماد}.$$

(پ) اگر نماد y_2 دریافت شود آن‌گاه اگر نماد x_1 یا x_3 ارسال شده باشد یک خطا انجام شده است. احتمال این خطا برابر است با

$$p(e|y_2)=p(x_1|y_2)+p(x_3|y_2)=1-p(x_2|y_2).$$

با کمک فرمول بیز مستقیماً نتیجه می‌شود که

$$p(x_2|y_2)=\frac{q(y_2|x_2)\cdot p(x_2)}{q(y_2)}=\frac{0{,}3\times\frac{1}{3}}{\frac{1}{2}}=\frac{1}{5},$$

بنابراین

$$p(e|y_2) = 1 - \frac{1}{5} = \frac{4}{5}.$$

(ت) از نقطه‌نظر گیرنده برای احتمال خطای p_e نتیجه می‌شود که

$$p_e = \sum_{j=1}^{3} q(y_j)\left[1 - p(x_j|y_j)\right].$$

با کمک فرمول بیز داریم

$$p(x_1|y_1) = \frac{q(y_1|x_1)\cdot p(x_1)}{q(y_1)} = \frac{0.5 \times \frac{1}{3}}{\frac{1}{3}} = \frac{1}{2},$$

$$p(x_2|y_2) = \frac{1}{5},$$

$$p(x_3|y_3) = 0,$$

بنابراین داریم

$$P_e = \frac{1}{3}(1 - \frac{1}{2}) + \frac{1}{2}(1 - \frac{1}{5}) + \frac{1}{6}(1 - 0) = \frac{1}{6} + \frac{2}{5} + \frac{1}{6} = \frac{11}{15}.$$

(ث) برای متوسط احتمال خطا از نقطه نظر فرستنده نتیجه می‌شود که

$$P_e = \sum_{i=1}^{3} p(x_i)\left[1 - q(y_i|x_i)\right] = \frac{1}{3}(1 - 0.5) + \frac{1}{3}(1 - 0.3) + \frac{1}{3}(1 - 0)$$
$$= \frac{1}{6} + \frac{7}{30} + \frac{1}{3} = \frac{11}{15}.$$

جوابهای بند (ت) و (ث) یکی هستند. قضیهٔ (۳.۴) را ببینید.

(ج) احتمال انتقال $q(y_3|x_3) = 0$، به عبارت دیگر احتمال این که یک نماد x_3 ارسال شده به نماد صحیح y_3 در طرف دیگر برود، صفر است. از طرف دیگر، احتمال این که نماد x_2 ارسال شده به نماد نادرست y_3 برود زیاد است، یعنی ۰٫۹.

(چ) مقدار اطلاع ارسال شده از رابطهٔ زیر نتیجه می‌شود

$$\mathrm{R} = H(X) - H(X|Y) = H(Y) - H(Y|X) = 1.46 - 1.18$$
$$= 0.28 \text{ بیت / نماد}.$$

ایهام $H(X|Y)$ را نیز می‌توان با کمک این رابطه محاسبه کرد، یعنی

$$H(X|Y) = H(X) - \mathrm{R}.$$

برای $H(X)$ داریم

$$H(X) = \log ۳ = ۱٫۵۸ \text{ بیت}.$$

بنابراین

$$H(X|Y) = ۱٫۵۸ - ۰٫۲۸ = ۱٫۳۰ \text{ بیت / نماد}.$$

(ح) نابرابری فانو بیان می‌کند که

$$H(X|Y) \le H(P_e) + P_e \log(n-۱).$$

با محاسبهٔ جملهٔ اوّل سمت راست نتیجه می‌شود:

$$H(P_e) = -P_e \log P_e - (۱-P_e)\log(۱-P_e)$$

$$= -\frac{۱۱}{۱۵}\log\frac{۱۱}{۱۵} - \frac{۴}{۱۵}\log\frac{۴}{۱۵} = ۰٫۸۴.$$

جمله دوم سمت راست برابر است با

$$P_e \log(n-۱) = \frac{۱۱}{۱۵}\log(۳-۱) = \frac{۱۱}{۱۵} = ۰٫۷۳.$$

مجموع دو جملهٔ اخیر نتیجه می‌دهد

$$H(P_e) + P_e \log(n-۱) = ۰٫۸۴ + ۰٫۷۳ = ۱٫۵۷ > H(X|Y) = ۱٫۳۰.$$

از این رو نابرابری فانو برقرار می‌باشد.

۷.۴ (الف) کانال دارای ساختمانی است که در شکل (۴. ۱۰-الف) نشان داده شده است.

در این حالت برای $H(Y|X)$ به دست می‌آوریم

$$H(Y|X) = -\sum_{i=۱}^{۲}\sum_{j=۱}^{۳} p(x_i)q(y_j|x_i)\log q(y_j|x_i)$$

$$= -\alpha[(۱-p)\log(۱-p) + p\log p] - (۱-\alpha)[p\log p + (۱-p)\log(۱-p)]$$

$$= -p\log p - (۱-p)\log(۱-p).$$

(ب) برای این که قادر باشیم R را محاسبه کنیم ابتدا باید $H(Y)$ را تعیین کنیم. با

$$q(y_۱) = \alpha(۱-p),$$

$$q(y_۲) = \alpha p + (۱-\alpha)p = p,$$

$$q(y_3)=(1-\alpha)(1-p),$$

نتیجه می‌شود که

$$H(Y)=-\alpha(1-p)\log\alpha(1-p)-p\log p-(1-\alpha)(1-p)\log(1-\alpha)(1-p).$$

پس از بازنویسی مجدّد این عبارت نتیجه می‌شود

$$H(Y)=-(1-p)\alpha\log\alpha-(1-p)(1-\alpha)\log(1-\alpha)-p\log p-(1-p)\log(1-p).$$

بنابراین برای R به دست می‌آوریم:

$$\mathsf{R}=H(Y)-H(Y|X)=(1-p)\{-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)\}.$$

(پ) ظرفیت از رابطهٔ زیر به دست می‌آید:

$$C=\max_{\alpha}\mathsf{R}$$

اگر چه می‌توان مقدار بهینه α را از $\frac{d\mathsf{R}}{d\alpha}=0$ تعیین کرد، ولی می‌توان مستقیماً از عبارت R در بالا با تشخیص این کـــه داخـل ابـرو دقیقـاً $H(\alpha,1-\alpha)$ اسـت و بنابراین حدّاکثر آن ۱ است آن را به دست آورد. از این رو مستقیماً نتیجه می‌شود که

$$C=1-p.$$

۸.۴ (الف) ماتریس انتقال به صورت زیر است:

$$Q=\begin{bmatrix}1 & 0\\ p & 1-p\end{bmatrix}.$$

بنابراین برای احتمالهای $q(y_j)$ به دست می‌آوریم

$$q(y_1)=1-\alpha+\alpha p \quad ; \quad q(y_2)=\alpha(1-p)$$

در این صورت $H(Y)$ برابر است با

$$H(Y)=-(1-\alpha+\alpha p)\log(1-\alpha+\alpha p)-\alpha(1-p)\log\alpha(1-p).$$

(ب) برای محاسبهٔ R باید نخست $H(Y|X)$ را پیدا کنیم. داریم

$$\begin{aligned}H(Y|X)&=-\sum_{i=1}^{2}\sum_{j=1}^{2}p(x_1)q(y_j|x_i)\log q(y_j|x_i)\\&=-(1-\alpha)\times 1\log 1-\alpha p\log p+(1-\alpha)\times 0\log 0+\alpha(1-p)\log(1-p)\\&=-\alpha p\log p-\alpha(1-p)\log(1-p).\end{aligned}$$

بنابراین داریم

$$R = H(Y) - H(Y|X) = -(1-\alpha+\alpha p)\log(1-\alpha+\alpha p) - \alpha(1-p)\log\alpha(1-p)$$

$$+\alpha p\log p + \alpha(1-p)\log(1-p)$$

$$= -(1-\alpha+\alpha p)\log(1-\alpha+\alpha p) - \alpha(1-p)\log\alpha + \alpha p\log p.$$

(پ) برای تعیین ظرفیت از عبارتی که برای R به دست آورده‌ایم نسبت به α مشتق می‌گیریم، نتیجه می‌شود

$$\frac{dR}{d\alpha} = (1-p)\log(1-\alpha+\alpha p) - (1-\alpha+\alpha p)\frac{p-1}{(1-\alpha+\alpha p)}\log e$$

$$-(1-p)\log\alpha - \alpha(1-p)\frac{1}{\alpha}\log e + p\log p$$

$$= (1-p)\log(1-\alpha+\alpha p) - (1-p)\log\alpha + p\log p.$$

با قرار دادن $\frac{dR}{d\alpha} = 0$ سپس برای $\alpha = \alpha_0$ نتیجه می‌شود که

$$\log\frac{1-\alpha_0+\alpha_0 p}{\alpha_0} = -\frac{p}{1-p}\log p$$

و

$$1-\alpha_0+\alpha_0 p = \alpha_0 p^{p/(p-1)}$$

و

$$\alpha_0 = \frac{1}{1-p+p^{p/(p-1)}}.$$

از این رو برای ظرفیت به دست می‌آوریم

$$C = -(1-\alpha_0+\alpha_0 p)\log(1-\alpha_0+\alpha_0 p) - \alpha_0(1-p)\log\alpha_0 + \alpha_0 p\log p$$

با

$$\alpha_0 = \frac{1}{1-p+p^{p/(p-1)}}.$$

با بازنویسی مجدّد نتیجه می‌شود

$$C = -\log\alpha_0 - \frac{p}{p-1}\log p = \log(1-p+p^{p/(p-1)}) - \frac{p}{p-1}\log p.$$

۵

منبع اطلاع پیوسته

۱.۵ توابع چگالی احتمال

قبل از معرّفی اندازهٔ اطلاع برای حالت پیوسته نخست باید مفهوم ***چگالی احتمال*** را تا حدّی مورد توجّه قرار داد. در حالت گسسته ممکن است با نگاشت برآمدهای یک آزمایش به روی خطّ حقیقی و سپس با نسبت دادن احتمال به هر عدد روی این خط حقیقی یک متغیّر تصادفی به دست آورد. برای متغیّرهای تصادفی پیوسته تعداد برآمدها بی‌نهایت هستند. در این حالت پیشامدها را در جای خود باید تعریف نمود تا بتوان احتمال معیّنی به آنها نسبت داد. با نگاشت برآمدها به روی خطّ حقیقی توسط تابع داده شده‌ای، متغیّر تصادفی پیوسته به صورت دامنه‌ای که یک بازه (متناهی یا بی‌نهایت) روی خطّ حقیقی است تعریف می‌شود.

یافتن احتمال پیشامد معیّنی، یعنی یک مقدار معیّن از متغیّر تصادفی پیوسته، به سادگی متغیّر تصادفی گسسته نیست. چون تعداد بی‌نهایت مقدار وجود دارد، احتمال هر نقطه باید صفر باشد، در صورتی که مجموع آنها باید برابر یک باشد. بنابراین، ممکن است از احتمال این که یک مقدار در یک زیربازه خط حقیقی قرار دارد صحبت کرد. این متناظر است با آنچه که ما آن را پیشامد می‌نامیم. از این رو ممکن است برای ولتاژ بین V ۱۰- و V ۱۰+، بگوییم احتمال دقیقاً ۶ ولت صفر است، در حالی که احتمال این که ولتاژ بین ۵٫۹ ولت و ۶٫۱۲ ولت باشد ممکن است به خوبی برابر ۰٫۸ باشد و بنابراین بزرگ‌تر از صفر است. پیشامدهایی به شکل $(a < \mathbf{x} \leq b)$ را می‌توان با کمک توابع توزیع تجمّعی توصیف

نمود. ***تابع توزیع تجمّعی*** یک متغیّر تصادفی $\mathbf{x}$ را به صورت احتمال پیشامد $\{\mathbf{x} \le x\}$ تعریف می‌کنیم:

$$F(x) = P(\mathbf{x} \le x) \quad , \quad -\infty < x < +\infty. \tag{۱.۵}$$

این بدین معناست که $F(x)$ احتمالی را نشان می‌دهد که متغیّر تصادفی یک مقدار در مجموعه $[-\infty, x)$ اختیار کند.

پیشامد $\{\mathbf{x} \le x\}$ و احتمال آن وقتی x تغییر کند، تغییر می‌نماید. از این رو $F(x)$ تابعی از متغیّر x است. برای متغیّرهای پیوسته این تابع توزیع یک تابع پیوسته است.

براساس اصول نظریهٔ احتمال به طور واضح تابع توزیع تجمّعی دارای ویژگیهای زیر است:

(I) $0 \le F(x) \le 1$،

(II) $\lim_{x\to\infty} F(x) = 1$،

(III) $\lim_{x\to-\infty} F(x) = 0$،

(IV) $F(x)$ تابعی ناکاهشی از x است، یعنی اگر $a < b$ آن‌گاه $F(a) \le F(b)$.

در شکل (۱.۵) یک مثال از تابع توزیع تجمّعی داده شده است.

تابع چگالی احتمال (pdf) $\mathbf{x}$، اگر وجود داشته باشد، به صورت مشتق تابع تجمّعی $F(x)$ تعریف می‌شود:

$$p(x) = \frac{dF(x)}{dx}, \tag{۲.۵}$$

که در آن $F(x)$ باید مشتق‌پذیر باشد. یک مثال از تابع چگالی احتمال در شکل (۲.۵) داده شده است.

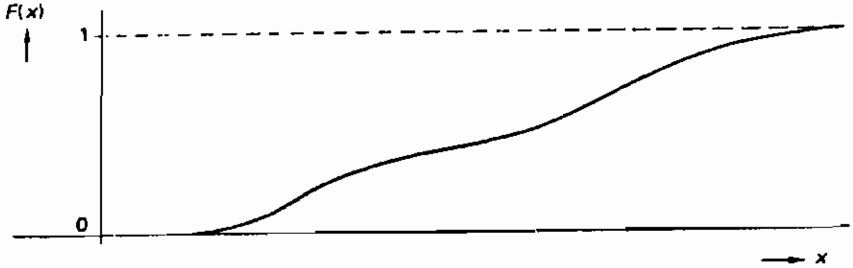
تعدادی از ویژگیهای تابع چگالی احتمال را خواهیم داد. چون تابع توزیع تجمّعی یکنوای ناکاهشی است (ویژگی (IV))، مستقیماً نتیجه می‌شود که

$$p(x) \ge 0.$$

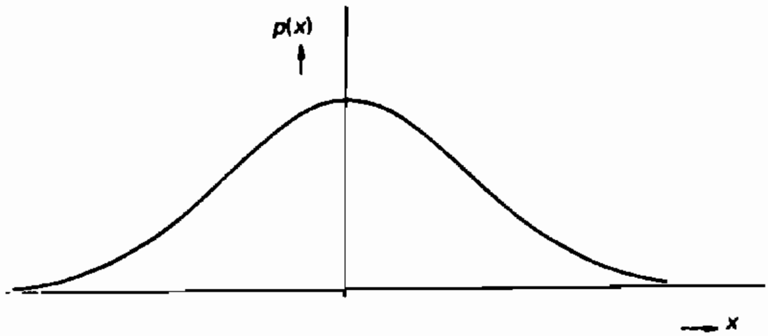
و همچنین

$$\int_{-\infty}^{x} p(u)\,du = F(x) - F(-\infty) = F(x). \tag{۳.۵}$$

از این رو می‌بینیم که مقدار تابع توزیع متناظر با سطح هاشور زده شده در شکل (۳.۵) می‌باشد. از این مستقیماً دیده می‌شود که



شکل ۱.۵- مثالی از تابع توزیع



شکل ۲.۵- مثالی از تابع چگالی احتمال

$$\int_{-\infty}^{\infty} p(x)\,dx = F(\infty) - F(-\infty) = ۱. \qquad (۴.۵)$$

بنابراین سطح زیر کلّ منحنی باید برابر یک باشد. اگر a و b را حدود انتگرال اختیار کنیم، در این صورت (شکل (۴.۵) را ببینید)

$$\int_{a}^{b} p(x)\,dx = F(b) - F(a) = P(\mathbf{x} \le b) - P(\mathbf{x} \le a) = P(a < \mathbf{x} \le b). \qquad (۵.۵)$$

از این رو می‌توان احتمال این‌که یک متغیّر تصادفی پیوسته مقداری بین a و b را اختیار کند با انتگرال‌گیری از چگالی احتمال روی (a,b) به دست آورد. اگر $a=b$ نتیجه می‌شود که $P(\mathbf{x}=a)=۰$ که با بیان ارائه شدهٔ قبلی که در حالت پیوسته احتمال یک مقدار خاص صفر است مطابقت دارد.

چگالی احتمال یک متغیّر تصادفی تقریباً همان نقشی را که احتمالها برای یک متغیّر تصادفی گسسته دارند انجام می‌دهد. برای یک متغیّر تصادفی پیوسته اگر چگالی احتمال روی فاصلهٔ معیّن (a,b) انتگرال‌گیری شده باشد تنها می‌توان از احتمال صحبت کرد. همچنین در این زمینه ممکن است چگالی احتمال را به صورت یک حد در نظر گرفت.

برای یک مقدار کوچک Δx نتیجه می‌شود که

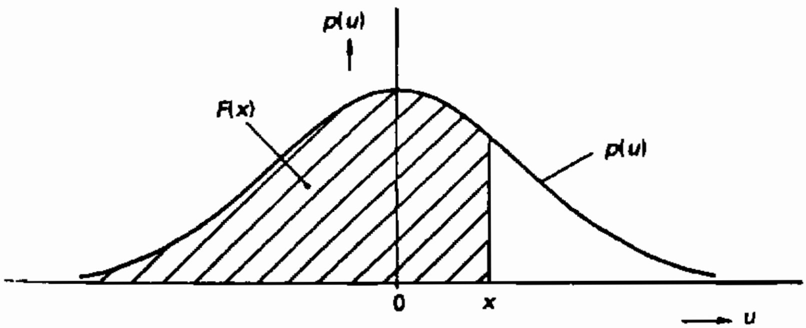
$$P(x < \mathbf{x} \le x + \Delta x) \approx p(x)\,\Delta x,$$

چون سطح زیر منحنی چگالی احتمال روی بازهٔ $(x, x + \Delta x)$ را می‌توان با مستطیلی با پهنای Δx تخمین زد. بنابراین

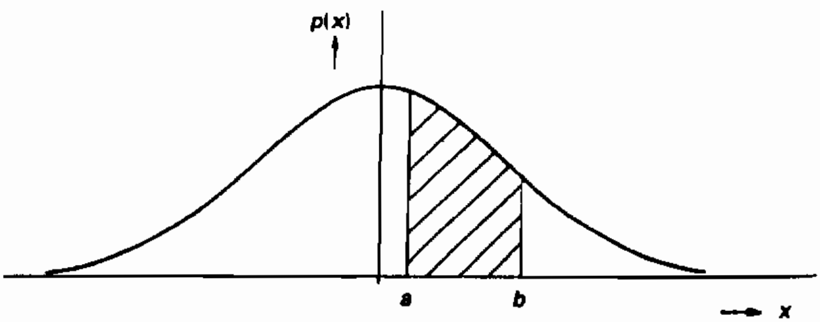
$$p(x) = \lim_{\Delta x \to 0} \frac{P(x < \mathbf{x} \le x + \Delta x)}{\Delta x}.$$

توجّه کنید که یک چگالی احتمال می‌تواند بزرگتر از یک باشد، که همچنین در مقایسه با احتمالها که حداکثر ممکن است برابر یک باشند یک اختلاف است. با وجود این، انتگرال $p(x)$ روی یک فاصله داده شده و از این رو احتمال این‌که $\mathbf{x}$ مقدار x را در این فاصله اختیار کند باید کمتر از یک باشد.

یک توزیع احتمال پیوسته مشهور توزیع یکنواخت است. یک متغیّر تصادفی پیوستهٔ $\mathbf{x}$ دارای توزیع یکنواخت است، اگر برای چگالی احتمال داشته باشیم:



شکل ۳.۵- رابطهٔ بین چگالی احتمال و تابع توزیع



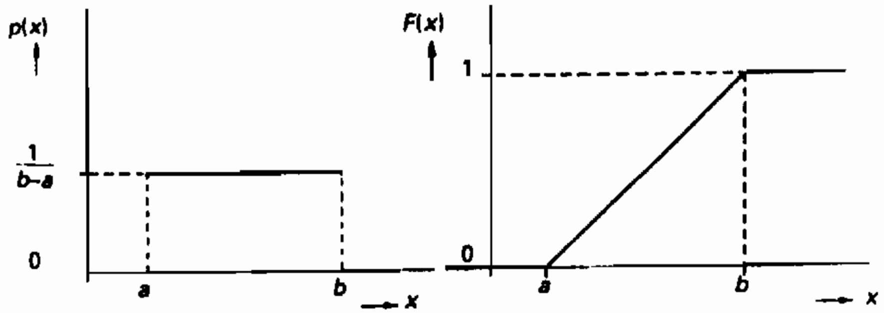
شکل ۴.۵

$$p(x)=\frac{۱}{b-a} \quad , \quad a\le x\le b, \qquad \qquad (۶.۵)$$
$$= ۰ \quad , \quad x<a, x>b.$$

تابع توزیع x عبارت است از

$$F(x)= ۰ \quad , \quad x<a,$$
$$=\frac{x-a}{b-a} \quad , \quad a\le x\le b, \qquad \qquad (۷.۵)$$
$$= ۱ \quad , \quad x>b.$$

توجّه کنید که برای $b-a<۱$ چگالی احتمال بزرگتر از یک است. توابع $p(x)$ و $F(x)$ در شکل (۵.۵) رسم شده‌اند.



شکل ۵.۵- توابع چگالی احتمال و توزیع برای توزیع یکنواخت

یک توزیع احتمال پیوستهٔ مشهور دیگر ***توزیع نرمال*** یا ***گاوسی*** است. این توزیع مهم است زیرا پدیده‌های فراوانی متغیّرهای تصادفی با توزیع نرمالند، نظیر نوفه در سیستمهای ارتباطی و اندازهٔ خطاهای انجام شده در زمان مشاهدهٔ سیستمها و سیگنالها. یک متغیّر تصادفی x دارای توزیع نرمال یا گاوسی است اگر چگالی احتمال آن به صورت زیر باشد:

$$p(x)=\frac{۱}{\sigma\sqrt{۲\pi}}\exp\left\{-\frac{(x-\mu)^{۲}}{۲\sigma^{۲}}\right\} \quad , \quad -\infty<x<\infty, \qquad (۸.۵)$$

که در آن μ و σ دو پارامتر هستند. برای توزیعهای پیوسته اغلب از ***میانگین*** یا ***امید*** x تعریف شده با

$$E(\mathbf{x})=\int_{-\infty}^{\infty} x\, p(x)\, dx, \qquad (۹.۵)$$

و ***واریانس*** استفاده می‌شود. واریانس اندازهٔ تغییرات مقادیر $\mathbf{x}$ حول میانگین آن می‌باشد و به صورت زیر تعریف می‌شود

$$\mathrm{var}(\mathbf{x}) = E[(\mathbf{x} - E(\mathbf{x}))^2] = \int_{-\infty}^{\infty} (x - E(\mathbf{x}))^2 \, p(x)\, dx. \qquad (۱۰.۵)$$

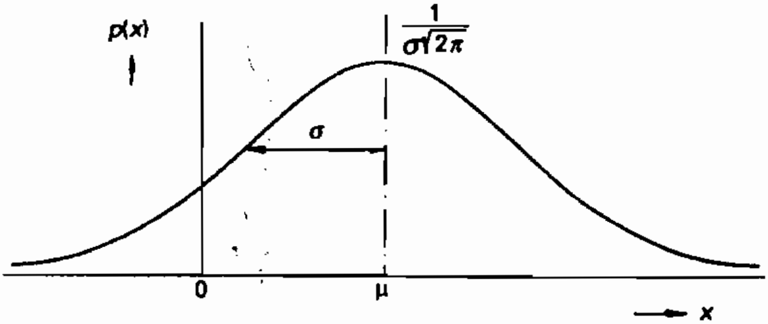
اگر $E(\mathbf{x})$ و $\mathrm{var}(\mathbf{x})$ را برای توزیع گاوسی محاسبه کنیم به دست می‌آوریم

$$E(\mathbf{x}) = \mu \qquad (۱۱.۵)$$

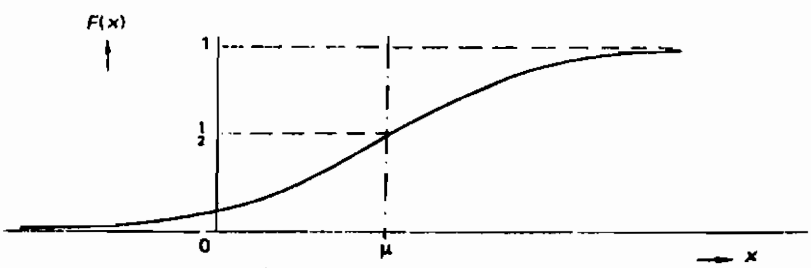
و

$$\mathrm{var}(\mathbf{x}) = \sigma^2. \qquad (۱۲.۵)$$

از این نتیجه می‌شود که پارامتر μ امید $\mathbf{x}$ و σ^2 واریانس را نشان می‌دهد. برای تکمیل مطلب باید یادآوری کرد که ریشهٔ دوم واریانس را ***انحراف معیار*** می‌نامند و از این رو در این حالت برابر σ است. پارامترهای μ و σ مشخصه‌هایی برای شکل $p(x)$ هستند. از این رو توزیع را نیز اغلب با $N(\mu,\sigma^2)$ نشان می‌دهند. پارامتر μ نقطهٔ تقارن نمودار $p(x)$ را تعیین می‌کند و $\mu-\sigma$ و $\mu+\sigma$ نقاط عطف این نمودارند (شکل (۶.۵) را ببینید).



شکل ۶.۵- توزیع گاوسی (تابع چگالی)



شکل ۷.۵- توزیع یک توزیع گاوسی

تابع توزیع متناظر عبارت است از (شکل (۷.۵) را ببینید)

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left\{-\frac{(u-\mu)^2}{2\sigma^2}\right\}du. \qquad (۱۳.۵)$$

جداولی برای تابع توزیع گاوسی $F(x)$ با $\mu=0$ و $\sigma=1$ وجود دارند.

یک توزیع گاوسی دلخواه $N(\mu,\sigma^2)$ را می‌توان به توزیع $N(0,1)$ با بـه کـار بـردن تبدیل $\mathbf{y}=\frac{\mathbf{x}-\mu}{\sigma}$ برای متغیّر تصادفی $\mathbf{x}$ استاندارد کـرد. توزیـع $N(0,1)$ را معمـولاً توزیـع نرمال استاندارد می‌نامند. از تقارن توزیع نتیجه می‌شود که $p(\mu+x)=p(\mu-x)$ و همچنین $F(\mu)=\frac{1}{2}$، چون در این صورت انتگرال‌گیری روی نصف دامنهٔ $\mathbf{x}$ انجام خواهد شد.

علاوه بر بررسی یک متغیّر تصادفی پیوسته همچنین ترکیبی از متغیّرهـای تصـادفی پیوسته را بررسی خواهیم کرد. فرض کنید که دو متغیّر پیوستهٔ $\mathbf{x}$ و $\mathbf{y}$ با چگالیهای احتمال $p(x)$ و $q(y)$ داریم؛ اکنون برای هر زوج عدد (x,y) می‌خواهیم $p(x,y)$ را بدانیم. بـرای انجام این کار ابتدا ***توزیع تجمّعی توأم*** برای دو متغیّر تصـادفی پیوسـته $\mathbf{x}$ و $\mathbf{y}$ را بررسـی می‌کنیم. این توزیع به صورت زیر تعریف می‌شود

$$F(x,y)=P(\mathbf{x}\le x,\mathbf{y}\le y). \qquad (۱۴.۵)$$

اکنون $p(x,y)$ ***چگالی احتمال دوبعدی*** یا ***توأم*** را به صورت مشـتق جزئـی نسـبت به x و y تعریف می‌کنیم:

$$p(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}. \qquad (۱۵.۵)$$

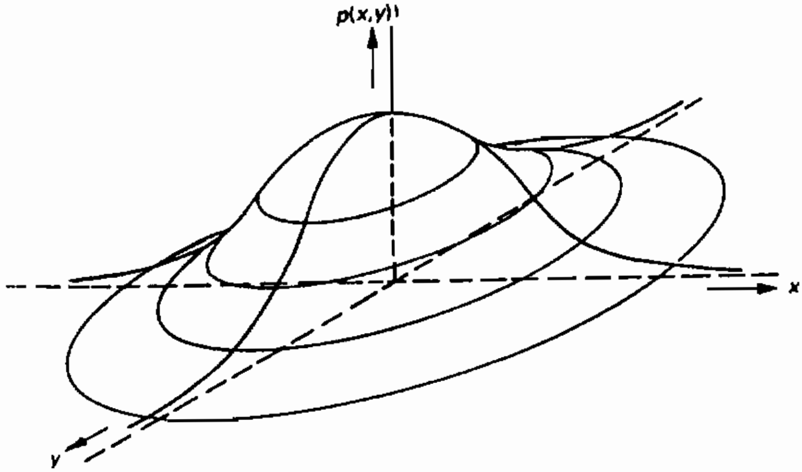
یک مثال از چگالی احتمال دوبعدی در شکل (۸.۵) داده شده است. هنوز هم درست است که

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\,dx\,dy=1, \qquad (۱۶.۵)$$

که به این معناست که حجم تحت تابع چگالی احتمال توأم باید برابر یک باشـد. درسـت مانند رابطهٔ بین احتمال حاشیه‌ای و توأم که در حالت گسسته وجود داشت اکنون داریم

$$p(x)=\int_{-\infty}^{\infty}p(x,y)\,dy, \qquad (۱۷.۵)$$

$$q(y)=\int_{-\infty}^{\infty}p(x,y)\,dx. \qquad (۱۸.۵)$$



شکل ۸.۵- مثالی از تابع چگالی احتمال دوبعدی

به عنوان مثالی از تابع چگالی احتمال دوبعدی توزیع گاوسی دوبعدی را در نظر بگیرید که به صورت زیر بیان می‌شود

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)\cdot(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}, \qquad (۱۹.۵)$$

با پارامترهای μ_1، μ_2، σ_1، σ_2 و ρ. در این جا ρ ***ضریب همبستگی*** است که $-1 \le \rho \le 1$ و در مورد آن بعداً بیشتر گفته خواهد شد. چگالی احتمال برای حالت $\mu_1 = \mu_2 = 0$، $\sigma_1 = \sigma_2$ و $\rho = 0$ در شکل (۸.۵) رسم شده است.

چگالیهای احتمال حاشیه‌ای متناظر با $p(x)$ و $q(y)$، که با به کار بردن معادلات (۱۷.۵) و (۱۸.۵) به دست آمده‌اند نیز، به ترتیب دارای توزیع نرمال $N(\mu_1, \sigma_1^2)$ و $N(\mu_2, \sigma_2^2)$ می‌باشند.

برای متغیّرهای تصادفی پیوسته که به یکدیگر وابسته می‌باشند می‌توان ***چگالیهای احتمال شرطی*** را به کار برد. کاربرد چگالیهای احتمال شرطی برای مسائلی در الکترونیک و تکنولوژی اطلاعات خیلی مهم است. این بدین علّت است که سیگنالهای نوفه دارای دامنهٔ پیوسته‌اند، و آن بدین معناست که مسائلی نظیر جدا کردن اطلاع حاصل از سیگنالها و سیگنالهای نوفه و تفسیر اطلاعات آلوده شده با نوفه طبیعتاً برحسب چگالی احتمال (یا

احتمال) شرطی فرمول‌بندی شده‌اند که در آن شرط توسط نتایج اندازه‌گیری شده ساخته می‌شود. تعریف تابع چگالی احتمال شرطی با مقایسه با حالت گسسته برحسب چگالی احتمال حاشیه‌ای و توأم فرمول‌بندی شده است. داریم

$$p(x|y) = \frac{p(x,y)}{q(y)}. \qquad (۲۰.۵)$$

رابطهٔ بین تابع چگالی احتمال حاشیه‌ای $p(x)$ و چگالی احتمال شرطی $p(x|y)$ به صورت زیر می‌باشد

$$p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy = \int_{-\infty}^{\infty} q(y)\,p(x|y)\,dy. \qquad (۲۱.۵)$$

فرمول بیز را نیز می‌توان با چگالیهای احتمال به صورت زیر بیان کرد

$$p(x|y) = \frac{p(x)\,p(y|x)}{q(y)}, \qquad (۲۲.۵)$$

یا همچنین

$$p(x|y) = \frac{p(x)\,p(y|x)}{\int_{-\infty}^{\infty} p(x)\,p(y|x)\,dx}. \qquad (۲۳.۵)$$

برای تکمیل مطلب باید اضافه کرد که سه رابطهٔ اخیر اگر x یک متغیّر پیوسته و y یک متغیّر گسسته باشد نیز برقرارند، که در این حالت این عبارات به شکل زیر در می‌آیند:

$$p(x) = \sum_i q(y_i)\,p(x|y_i),$$

$$p(x|y_i) = \frac{p(x)\,q(y_i|x)}{q(y_i)},$$

$$p(x|y_i) = \frac{p(x)\,q(y_i|x)}{\int_{-\infty}^{\infty} p(x)\,q(y_i|x)\,dx}.$$

این روابط در نظریهٔ اطلاع و به خصوص در نظریهٔ ارزیابی آماری به کار برده می‌شوند. در مقایسه با حالت گسسته، استقلال آماری دو متغیّر تصادفی پیوسته x و y را می‌توان به صورت زیر تعریف کرد

$$p(x,y) = p(x)\,q(y). \qquad (۲۴.۵)$$

در این حالت روابط زیر نیز درست می‌باشند:

$$q(y|x) = q(y),$$

$$p(x|y) = p(x).$$

با به کار بردن احتمالهای شرطی و توابع چگالی احتمال شرطی می‌توان وجود وابستگی بین دو متغیّر تصادفی پیوسته را نشان داد. اغلب به جای آن ***کوواریانس*** یا **همبستگی** نیز به کار برده می‌شود. فرض کنید $\mathbf{x}$ و $\mathbf{y}$ دو متغیّر تصادفی پیوسته باشند؛ در این صورت ***کوواریانس*** به صورت زیر تعریف می‌شود

$$\begin{aligned}\mathrm{cov}(\mathbf{x},\mathbf{y}) &= E[(\mathbf{x}-E(\mathbf{x}))\cdot(\mathbf{y}-E(\mathbf{y}))]\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E(\mathbf{x}))(y-E(\mathbf{y}))\,p(x,y)\,dx\,dy. \qquad (۲۵.۵)\end{aligned}$$

توجّه کنید که کوواریانس متغیّر تصادفی $\mathbf{x}$ با خودش، برابر است با

$$\mathrm{cov}(\mathbf{x},\mathbf{x}) = E[(\mathbf{x}-E(\mathbf{x}))^{2}] = \mathrm{var}(\mathbf{x}). \qquad (۲۶.۵)$$

اگر کوواریانس را نسبت به واریانسهای $\mathbf{x}$ و $\mathbf{y}$ استاندارد کنیم در این صورت **ضریب همبستگی** ρ به دست می‌آید، که قبلاً در عبارت توزیع گاوسی دوبعدی با آن مواجه شدیم. ضریب همبستگی به صورت زیر می‌باشد

$$\rho = \frac{\mathrm{cov}(\mathbf{x},\mathbf{y})}{\sqrt{\mathrm{var}(\mathbf{x})\,\mathrm{var}(\mathbf{y})}}. \qquad (۲۷.۵)$$

می‌توان نشان داد که $|\rho| \le 1$.

همبستگی بین $\mathbf{x}$ و $\mathbf{y}$ به صورت زیر تعریف می‌شود

$$R(\mathbf{x},\mathbf{y}) = E(\mathbf{x}\cdot\mathbf{y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\,p(x,y)\,dx\,dy. \qquad (۲۸.۵)$$

رابطهٔ بین کوواریانس و همبستگی را با کمک معادلهٔ (۲۵.۵) می‌توان به دست آورد:

$$\begin{aligned}\mathrm{cov}(\mathbf{x},\mathbf{y}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{xy - x\,E(\mathbf{y}) - y\,E(\mathbf{x}) + E(\mathbf{x})E(\mathbf{y})\}\,p(x,y)\,dx\,dy\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\,p(x,y)\,dx\,dy - E(\mathbf{x})E(\mathbf{y}) - E(\mathbf{y})E(\mathbf{x}) + E(\mathbf{x})E(\mathbf{y})\\ &= R(\mathbf{x},\mathbf{y}) - E(\mathbf{x})E(\mathbf{y}). \qquad (۲۹.۵)\end{aligned}$$

۲.۵ سیگنالهای تصادفی

سیگنالهایی را که دامنهٔ نوسانات آنها در نقطهٔ دلخواهی از زمان وابسته به زمان است، ***سیگنالهای تصادفی*** می‌نامیم. معمولاً یک سیگنال تصادفی را به صورت $\{\mathbf{x}(t), t \in T\}$ تعریف می‌کنیم که در آن T می‌تواند یک فاصلهٔ زمانی معیّنی، بلکه همچنین یک مجموعهٔ نقاطی از زمان (برای مثال شروع هر ساعت) باشد. منظور ما از $\mathbf{x}(t)$ مقدار سیگنال در نقطه زمانی t است.

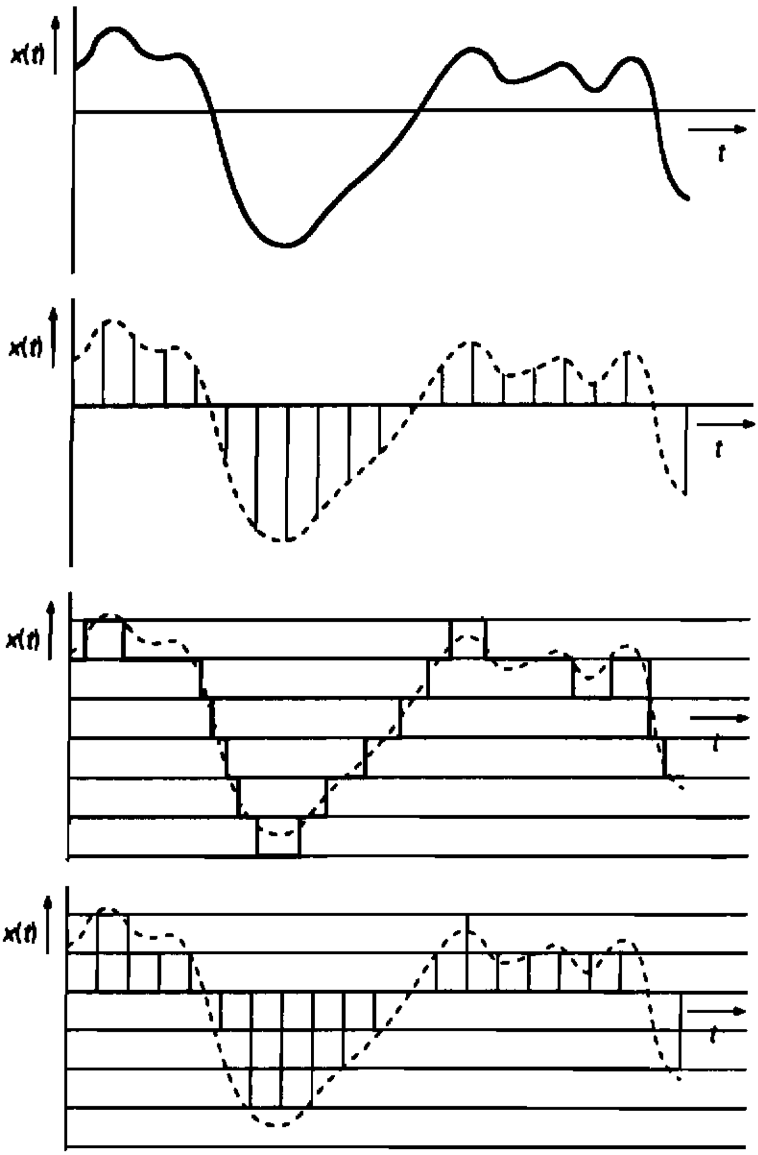
در نقطهٔ داده شده $t_\circ$، $\mathbf{x}(t_\circ)$ یک متغیّر تصادفی است که مقداری را که سیگنال می‌تواند در زمان $t_\circ$ اختیار کند تعیین می‌کند و فرض می‌شود که احتمال رخ دادن آن معلوم است. چون سیگنال $\mathbf{x}(t)$ برای تمام tهای متعلق به T داده شده است، این در اصل تعیین می‌کند که سیگنال $\mathbf{x}(t)$ چه شکلی در کل می‌تواند اختیار کند و چقدر محتمل است که آنها رخ دهند. از این رو یک سیگنال تصادفی $\mathbf{x}(t)$ را برای هر نقطهٔ از زمان $t \in T$ به عنوان متغیّر تصادفی می‌توان در نظر گرفت. از طرف دیگر، می‌توان همچنین تعیین کرد که سیگنال در هر زمان برای تمام نقاط زمانی در مجموعهٔ T شبیه چیست. در این صورت مقادیر به دست آمدهٔ سیگنال را به طور کامل در نظر می‌گیریم. در این صورت از مسیر سیگنال تصادفی که با $x(t)$ نشان داده می‌شود صحبت می‌کنیم. یک مسیر از یک سیگنال تصادفی را می‌توان به صورت نتیجهٔ یک آزمایش در نظر گرفت: سیگنال تصادفی $\{\mathbf{x}(t), t \in T\}$ را یک بار بیرون آورده و نتیجه را مشاهده می‌کنیم (یعنی، نتیجهٔ مقادیر سیگنال $[x(t), t \in T]$). مجموعهٔ همهٔ مسیرهای ممکن اغلب به صورت ***دسته*** نشان داده می‌شود.

با سیگنالهای مشابه، مقدار سیگنال در هر نقطهٔ زمان یک کمیت پیوسته است، به طوری که فرایند تصادفی $\{\mathbf{x}(t), t \in T\}$ برای یک نقطهٔ زمانی t داده شده یک متغیّر تصادفی پیوسته با یک چگالی احتمال معیّن $p(x)$ است. همچنین سیگنال، پدیده‌ای پیوسته به صورت تابعی از زمان می‌باشد، که بدین معناست که مجموعهٔ T یک فاصلهٔ زمانی معیّن است. این زمان اغلب زمان اندازه‌گیری یا زمان مشاهده است لکن می‌تواند زمان واقعی که سیگنال رخ می‌دهد نیز باشد.

به طور کلّی چهار نوع سیگنال تصادفی وجود دارند که در شکل (۹.۵) خلاصه شده‌اند.

در مطالب ذکر شده در بالا به صراحت فرض شده است که سیگنالها، سیگنالهای یک‌بعدی به عنوان تابعی از زمان می‌باشند. با وجود این، همچنین می‌توان سیگنالهای دو یا حتّی بیش از دوبعدی مثلاً تصاویر تلویزیونی یا ثبّاتهای چندکانالی سیگنالهای زلزله‌ای را

در نظر گرفت. همچنین می‌توان شدّت یک تصویر تلویزیونی را به عنوان یک فرایند تصادفی در نظر گرفت که در آن مختصّات (x,y) نقاط نوری روی صفحه نقش مشابهی را با پارامتر زمان انجام می‌دهند.



شکل ۹.۵- انواع سیگنالهای تصادفی

اگر چه یک سیگنال مشابه دارای یک ویژگی پیوسته نسبت به هر دو مقادیر

سیگنال و زمان به عنوان یک پارامتر است، این توصیف برای تابع بودن خیلی پیچیده است. در عمل اغلب خود را به سیگنالهای زمان گسسته که تنها در یک نقطهٔ زمانی معیّن معلوم است محدود می‌کنیم.

رابطهٔ بین پهنای باند و فاصلهٔ زمان بین نقاط متوالی زمان که در آن یک سیگنال بایستی تعیین شود توسط ***قضیهٔ نمونه‌گیری*** داده شده است.

قضیهٔ ۱.۵ (قضیهٔ نمونه‌گیری در دامنهٔ زمان)

اگر یک سیگنال $x(t)$ دارای پهنای باند W هرتز باشد، یعنی، W سیکل بر ثانیه، در این صورت سیگنال با معلوم بودن مختصّات آن در یک سری از نقاط که با فاصلهٔ $\frac{۱}{۲W}$ ثانیه از هم قرار دارند کاملاً تعیین می‌شود، سری در سراسر طول دامنهٔ زمان گسترش می‌یابد.

سیگنال پیوسته $x(t)$ را می‌توان از نمونه‌های $x(\frac{k}{۲W})$ به صورت زیر بازسازی کرد:

$$x(t)=\sum_{k=۱}^{۲WT} x(\frac{k}{۲W})\frac{\sin ۲\pi W(t-\frac{k}{۲W})}{۲\pi W(t-\frac{k}{۲W})}. \qquad (۳۰.۵)$$

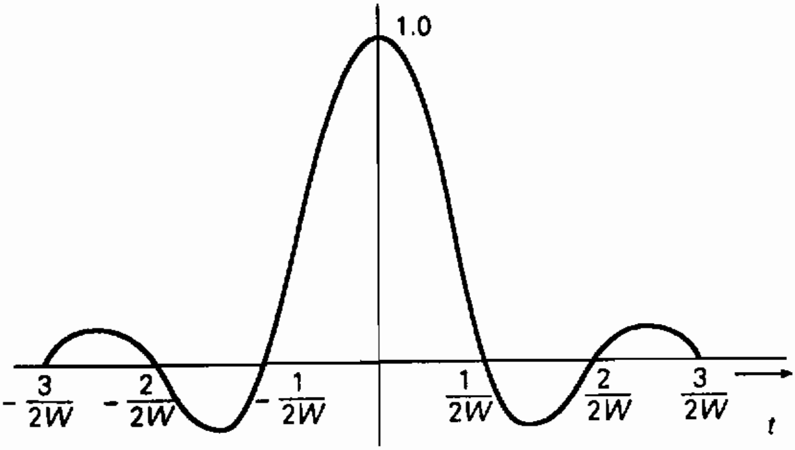
■

اثبات قضیه در این جا ارائه نخواهد شد. تابع $\frac{\sin ۲\pi Wt}{۲\pi Wt}$ که همچنین با $\text{sinc}\,۲\pi Wt$ نشان داده می‌شود دارای ویژگیهایی است که مقدارش در $t=۰$ یک است و در $t=\frac{k}{۲W}$، یعنی همهٔ نقاط نمونه به استثنای $t=۰$ برابر صفر است (شکل (۵. ۱۰) را ببینید).

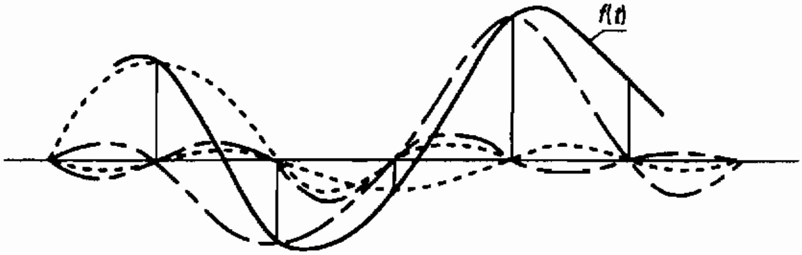
سیگنال اصلی $x(t)$ را می‌توان از مقادیر نقطهٔ نمونه با به کار بردن تابع سینوس‌وار در هر نقطه نمونه ضرب‌در $x(\frac{k}{۲W})$ بازسازی کرد. با افزودن همهٔ توابع سینوس‌وار به یکدیگر $x(t)$ به دست می‌آید. شکل (۱۱.۵) را ببینید.

از این رو، برای سیگنالهای با باند محدود یک نمایش زمان گسسته کافی است، حتّی اگر آنها در اصل زمان-پیوسته باشند. با توجّه به نمونه‌های یک سیگنال تصادفی زمان-گسسته که با $x(\frac{k}{۲W})$ نشان داده شده است، اغلب برای سادگی با $۲W=۱$ مرتّب می‌شوند به طوری که می‌توان نماد $x(k)$، $k=۰,۱,۲,\ldots$ را به کار برد. به طور کلّی، اگر مدّت سیگنال T باشد این مدّت در $۲WT$ نمونه سیگنال موجود نتیجه خواهد شد. با در نظر گرفتن سیگنال گسسته-زمان آنالیز ساده‌تر می‌شود بدون این‌که از کلّیّت نتایج کاسته شود.

با بازگشت به سیگنالهای تصادفی $\{\mathbf{x}(t), t\in T\}$، پس از به کار بردن قضیهٔ



شکل ۵. ۱۰- تابع سینوس‌وار



شکل ۵.۱۱- توضیح قضیهٔ نمونه‌گیری

نمونه‌گیری، می‌توان یک سیگنال را به صورت یک سری از نمونه‌های (گسسته زمان) که در آن مقدار هر نمونه را می‌توان به عنوان یک متغیّر تصادفی در نظر گرفت نمایش داد. برای تعیین سیگنال تصادفی کامل باید چگالی احتمال توأم $p(\mathbf{x}) = p(x(t_1), x(t_2), \ldots, x(t_N))$ را برای کلّ مجموعهٔ N لحظه نمونه‌گیری به دست آورد. در این صورت هر «مقدار» $\mathbf{x}$ در واقع مسیری از فرایند تصادفی $\{\mathbf{x}(t), t \in T\}$ است که در لحظات نمونه‌گیری t_1، t_2، ...، t_N تعیین شده است.

اغلب غیرضروری یا حتّی غیرممکن است که چگالی احتمال توأم همهٔ نمونه‌ها را به کار بریم. اگر خود را به توصیف مرتبه-اوّل، یعنی با $N = ۱$ محدود کنیم، در این صورت می‌توانیم مشخّصات سیگنال را برای هر نمونه بررسی کنیم. چون وابستگی بین نمونه‌ها در عمل نقش بزرگی را بازی می‌کند، معمولاً یک توصیف مرتبه-دوم یعنی $N = ۲$ پسندیده یا مورد لزوم است. این معمولاً کافی نیز هست. یک سیگنال تصادفی مرتبه-دوم را ممکن

است با $\{\mathbf{x}(t_i,t_j)\ ;\ t_i,t_j \in T\}$ یا به عبارت دیگر از طریق دو نمونه با چگالی احتمال توأم $p(x(t_i),x(t_j))$ تعیین کرد.

در اصل سیگنال را برای چگالی احتمال $p(x(t_i),x(t_j))$ داده شده می‌توان تجزیه و تحلیل کرد. با وجود این، اغلب کمیّتهایی نظیر میانگین، همبستگی یا کوواریانس بررسی می‌شوند نه خود چگالی احتمال.

تابع خودهمبستگی به صورت زیر تعریف می‌شود

$$R_{xx}(t_i,t_j) = E\{\mathbf{x}(t_i)\cdot\mathbf{x}(t_j)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t_i)x(t_j)p(x(t_i),x(t_j))dx(t_i)dx(t_j). \quad (۳۱.۵)$$

و از این رو به t_i و t_j بستگی دارد.

اکنون اگر لحظات نمونه‌گیری t_i و t_j را به کار بریم در این صورت می‌توان مقادیر تابع خودهمبستگی را با کمک یک ماتریس به نام ***ماتریس خودهمبستگی*** $\mathbf{R}_{xx}$ نیز تعیین کرد:

$$\mathbf{R}_{xx} = \begin{bmatrix} R_{11} & \cdots & R_{1N} \\ \vdots & & \vdots \\ R_{N1} & \cdots & R_{NN} \end{bmatrix}, \quad (۳۲.۵)$$

که در آن $R_{ij} = R_{xx}(t_i,t_j)$.

این تعریف هنوز هم خیلی کلّی است چون همبستگی بین دو نمونه $x(t_i)$ و $x(t_j)$ ممکن است برای زوج نقاط زمان مختلف t_i و t_j متفاوت باشد. چنین سیگنالهایی را ***نا-مانا*** می‌نامند. با وجود این، اغلب می‌توان فرض کرد که سیگنالها ***مانا*** هستند که در آن کوواریانس تنها به اختلاف زمان $\tau = t_i - t_j$ بستگی دارد و به لحظات زمان مطلق t_i و t_j بستگی ندارد. در این صورت آن را ***سیگنال مانای ضعیف*** می‌نامیم که برای آن داریم

$$R_{xx}(t_i,t_j) = R_{xx}(t_i - t_j) = R_{xx}(\tau) = E\{\mathbf{x}(t_i)\mathbf{x}(t_i - \tau)\}. \quad (۳۳.۵)$$

برای $\tau = ۰$ نتیجه می‌شود $R_{xx(۰)} = E\{\mathbf{x}(t_i)^۲\}$ که برابر ***متوسط توان*** برای هر نمونه است و همچنین با P_x نشان داده می‌شود.

در حالت سیگنال مانای ضعیف ممکن است چگالی احتمال $p(x(t_i),x(t_j))$ یک ***سیگنال مانای ضعیف*** هنوز هم به نقاط زمان مطلق t_i و t_j وابسته باشد. بنابراین جملهٔ ***اکیداً مانا*** نیز برای نشان دادن این که چگالی احتمال $p(x(t_i),x(t_j))$ خودش مانای زمان است، یعنی تنها به $t_i - t_j$ بستگی دارد، به کار برده می‌شود. یک سیگنال اکیداً مانا یک سیگنال مانای ضعیف نیز هست ولی عکس آن ***ذاتاً*** درست نمی‌باشد.

علاوه بر خودهمبستگی اغلب از ***اتوکوواریانس*** نیز استفاده می‌شود. این به صورت زیر تعریف می‌شود

$$K_{xx}(t_i,t_j) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x(t_i)-E\{\mathbf{x}(t_i)\})\cdot(x(t_j)-E\{\mathbf{x}(t_j)\})\,p(x(t_i),x(t_j))\,dx(t_i)\,dx(t_j), \tag{۳۴.۵}$$

اختلافش با همبستگی صرفاً در این است که اکنون امید $x(t_i)$ و $x(t_j)$ از $x(t_i)$ و $x(t_j)$ کم شده است. برای کاربردهای زیادی این میانگینها صفر هستند، به طوری که همبستگی و کوواریانس در این صورت با یکدیگر برابرند. ***ماتریس اتوکوواریانس*** برای N نمونه برابر است با

$$\mathbf{K}_{xx} = \begin{bmatrix} K_{11} & \cdots & K_{1N} \\ \vdots & & \vdots \\ K_{N1} & \cdots & K_{NN} \end{bmatrix}, \tag{۳۵.۵}$$

که در آن $K_{ij} = K_{xx}(t_i,t_j)$. به سادگی دیده می‌شود که ماتریس خودهمبستگی و ماتریس اتوکوواریانس متقارن هستند، یعنی $R_{ij} = R_{ji}$ و $K_{ij} = K_{ji}$.

یک دسته از سیگنالهای مهم که برای آنها کافی است که سیگنال، سیگنال مانای ضعیف باشد با سیگنالهای تصادفی گاوسی تشکیل می‌شوند. برای این سیگنالها ماتریس اتوکوواریانس یک توصیف کافی از سیگنال است.

سیگنال تصادفی $\{\mathbf{x}(t), t\in T\}$ را ***سیگنال گاوسی*** نامیم اگر همهٔ چگالیهای احتمال N بعدی آن $p(x(t_1),\ldots,x(t_N))$ برای $N=1,2,\ldots$ چگالیهای احتمال گاوسی N بعدی باشند. چگالی احتمال گاوسی برای $N=1$، بدین ترتیب برای یک نمونه، در معادلهٔ (۸.۵) داده شده است. برای $N=2$ توزیع گاوسی دوبعدی در معادلهٔ (۱۹.۵) داده شده است.

در حالت کلّی توزیع گاوسی N بعدی ساده است که از نماد بُرداری استفاده کنیم. در این صورت عبارت برای $\tilde{\mathbf{x}}(x(t_1),\ldots,x(t_N))$ به صورت زیر است.

$$p(\tilde{\mathbf{x}}) = \frac{1}{(2\pi)^{N/2}|\mathbf{K}_{xx}|^{1/2}}\exp\{-\frac{1}{2}(\tilde{x}-\tilde{\mu})\mathbf{K}_{xx}^{-1}(\tilde{x}-\tilde{\mu})^T\}, \tag{۳۶.۵}$$

که در آن $\tilde{\mu}$ بردار مقادیر میانگین و $|\mathbf{K}_{xx}|$ دترمینان ماتریس اتوکوواریانس $\mathbf{K}_{xx}$ می‌باشد.

سیگنال گاوسی تنها سیگنالی است که با میانگینش و ماتریس اتوکوواریانس تعیین می‌گردد.

اگر فرایند مانا باشد کوواریانس K_{ij} تنها به اختلاف زمان $|t_i - t_j|$ بستگی دارد. برای

مثال این بدین معناست که $K_{ij}=K_1$ برای تمام زمان t_i و t_j به طوری که $t_j=t_i+1$ یا $t_j=t_i-1$. در این صورت ماتریس اتوکوواریانس به صورت زیر است

$$\mathbf{K}_{xx}=\begin{bmatrix} K_0 & K_1 & \cdots & \cdots & K_{N-1} \\ K_1 & K_0 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & K_1 \\ K_{N-1} & \cdots & \cdots & K_1 & K_0 \end{bmatrix}, \tag{۳۷.۵}$$

و بنابراین یک ساختار بسیار خاصّی را نشان می‌دهد. ماتریس به این شکل را ***ماتریس تئوپولی***[1] می‌نامند، ردیف اوّل این ماتریس مقادیر نمونه‌گیری شدهٔ تابع اتوکوواریانس فرایند گاوسی را می‌دهد.

سیگنال گاوسی چندین ویژگی دارد که کاربرد آن را جالب می‌سازد. احتمالاً مهمترین ویژگی آن است که هر ترکیب خطی که روی سیگنال گاوسی انجام شود همچنین به یک سیگنال گاوسی منجر می‌شود. به این علّت، فرایند گاوسی نقشی با تجزیه و تحلیل سیگنالهای تصادفی بازی می‌کند که با نقش سیستمهای خطّی در نظریهٔ سیستمها قابل مقایسه است.

توصیف سیگنالهای تصادفی تا این‌جا محدود به دامنهٔ زمان باقی‌مانده است. با وجود این، در این جا می‌توان از دامنهٔ فرکانس نیز استفاده نمود. بنابراین ***توان چگالی طیفی*** بسیار مهم است؛ از این رو که اندازه‌ای برای مقدار توان در پهنای باند سیگنال تصادفی است.

اغلب این ***توان چگالی طیفی***، $S_{xx}(\omega)$، را که به طور ساده طیف می‌نامیم، می‌توان آن را از تابع اتوکوواریانس با کمک تبدیل فوریه به دست آورد. آن را به صورت زیر تعریف می‌کنیم

$$S_{xx}(\omega)=\int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j\omega\tau}\,d\tau, \tag{۳۸.۵}$$

$$R_{xx}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\omega)e^{+j\omega\tau}\,d\omega. \tag{۳۹.۵}$$

برای حالت $\tau=0$ داریم

$$R_{xx}(0)=\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\omega)d\omega=P_x. \tag{۴۰.۵}$$

این بدین معناست که همان‌طور که قبلاً دیده‌ایم متوسط توان یک سیگنال را می‌توان

1. Toeplitz

از تابع اتوکوواریانس با قرار دادن $\tau = ۰$ به دست آورد؛ همچنین با انتگرال‌گیری از تــوان چگالی طیفی روی کلّ دامنهٔ فرکانس می‌توان آن را به دست آورد.

۳.۵ اندازهٔ اطلاع پیوسته

اندازهٔ اطلاع پیوسته را می‌توان براساس اطلاع گسسته به صورت زیر به دســت آورد. همان‌طور که در بالا دیدیم یک تابع چگالی احتمــال پیوســته را می‌تــوان بــا چگالیهــای احتمالی که روی فاصله‌هایی به طول Δx ثابتند تخمین زد. فرض کنید که p_i مقدار ثــابت احتمال p_i در فاصلهٔ i باشد. برای تضمین این‌که $\sum_i p_i = ۱$، قــرار می‌دهیــم $p_i = p(x_i)\Delta x$ که در آن x_i نقطه‌ای در فاصلهٔ i است به قسمی‌که $p(x_i)\Delta x$ برابر با سطح زیر تابع چگالی احتمال پیوسته $p(x)$ در فاصلهٔ i باشد. اکنون حالتی است که

$$\begin{aligned} H(X) &= -\sum_{i=۱}^{n} p_i \log p_i \\ &= -\sum_{i=۱}^{n} p(x_i)\Delta x \log p(x_i)\Delta x \\ &= -\sum_{i=۱}^{n} p(x_i)\Delta x \log p(x_i) - \log \Delta x. \end{aligned} \qquad (۴۱.۵)$$

با حد گرفتن وقتی که $\Delta x \to ۰$، داریم

$$\lim_{\Delta x \to ۰} H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) - \lim_{\Delta x \to ۰} \log \Delta x. \qquad (۴۲.۵)$$

روشن است که جملهٔ دوم بی‌نهایت خواهد شد. از این رو مقدار اطــلاع یــک متغــیّر تصادفی پیوسته همواره بی‌نهایت است. در واقع نتیجه تعجب‌آور نیست. با تفســیر کــردن اندازهٔ اطلاع به صورت متوسط تعداد جوابهای بلی یا خیر داده شدهٔ لازم برای تحلیل کردن عدم‌حتمیت، این عدد در حالت پیوسته بی‌نهایت خواهد بود.

با وجود این، این فقط یک دیدگاه نظری است. در عمل اندکی انــدازه عدم‌حتمیت وجود خواهد داشت.

این همچنین Δx را متناهی می‌کند. به طور کلّی Δx برابر واحد انتخاب شده است که توسط آن جملهٔ دوم برابر صفر می‌شود. این موضوع به تعریف زیر منجر می‌شود.

تعریف ۱.۵

برای متغیّر تصادفی پیوسته x با تابع چگالی احتمال $p(x)$ مقدار اطلاع برابر است با

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx. \qquad (۴۳.۵)$$

■

واضح است که تعریف اندازهٔ اطلاع پیوسته بر مبنای مشابهت با اطلاع گسسته است تا براساس بیان مشتق ریاضی.

به عنوان نتیجه $H(X)$ تعریف شده با این روش برای یک متغیّر تصادفی پیوسته می‌تواند منفی شود، که در تناقض با حالت گسسته است.

همچنین مهم است بدانیم که چه چگالی احتمالی برای متغیّر تصادفی پیوسته به ماکسیمم مقدار اطلاع منجر می‌شود. با وجود این، مشتق مغایر است با آنچه که برای حالت گسسته داده شده است. این به علّت این حقیقت است که معمولاً باید محدودیتهای اضافی برای متغیّرهای تصادفی پیوسته تحمیل گردد. برای مثال این محدودیتها ممکن است یک دامنهٔ نوسان کراندار یا یک توان (واریانس) ثابت باشد. از این رو طبیعت این محدودیتها توأماً طبیعت چگالی احتمال را که به ماکسیمم مقدار اطلاع منجر می‌شود تعیین می کنند. دو حالت یعنی محدود کردن دامنهٔ نوسان و محدود کردن توان (یا واریانس) بررسی خواهد شد. در قضیهٔ زیر در حالتی که دامنه بین $-A$ و $+A$ محدود شده باشد چگالی احتمالی تعیین خواهد شد که به ماکسیمم مقدار اطلاع منجر می‌شود.

قضیهٔ ۲.۵

برای سیگنالی با دامنهٔ نوسان محدود شده در دامنهٔ $(-A, +A)$، مقدار اطلاع $H(X)$ ماکسیمم است اگر و تنها اگر

$$p(x) = \frac{۱}{۲A}.$$

و ماکسیمم مقدار برابر است با

$$H(X) = \log ۲A.$$

برهان

برای حلّ این مسأله از روش محاسبهٔ تغییرات استفاده خواهد شد. هدف تعیین چگالی احتمال $p(x)$ است که برای آن

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx,$$

ماکسیمم باشد، که در آن $p(x)$، با دامنهٔ نوسان کراندار، باید در رابطهٔ زیر صدق کند

$$\int_{-A}^{A} p(x)\,dx = ۱.$$

برای پایان دادن به این مطلب از تابع

$$G(x) = -p(x)\log p(x) + \alpha p(x)$$

نسبت به $p(x)$ مشتق گرفته و سپس برابر صفر قرار می‌دهیم. نتیجه می‌شود

$$-\log p(x) - \log e + \alpha = ۰$$

یا

$$\ln p(x) = \frac{\alpha}{\log e} - ۱ = k$$

به قسمی که $p(x) = e^k$.

داریم

$$\int_{-A}^{A} p(x)\,dx = ۱,$$

بنابراین با جایگزینی $p(x)$ نتیجه می‌شود

$$\int_{-A}^{A} e^k\,dx = ۱ \;\Rightarrow\; [e^k x]_{-A}^{A} = e^k ۲A = ۱,$$

که نتیجه می‌دهد

$$p(x) = \frac{۱}{۲A}.$$

با جایگزینی $p(x) = \frac{۱}{۲A}$ در $H(X)$ به دست می‌آوریم

$$H(X) = -\int_{-A}^{A} \frac{۱}{۲A}\log\left(\frac{۱}{۲A}\right)dx = \log ۲A.$$

سرانجام در حالت کراندار بودن دامنهٔ نوسان، چگالی احتمال یکنواخت ماکسیمم مقدار اطلاع متناسب با ماکسیمم دامنهٔ نوسان را می‌دهد. این تا حدّی با حالت متغیّر تصادفی گسسته که دارای ماکسیمم مقدار اطلاع است، اگر دارای توزیع احتمال یکنواخت باشد، مطابقت دارد.

حالت مهم دیگر وقتی است که توان یک سیگنال کراندار باشد، که به تثبیت واریانس نمونه‌ها منجر می‌گردد.

قضیهٔ ۳.۵

برای یک سیگنال با توان ثابت $\sigma^۲$،

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 p(x) dx,$$

$H(X)$ ماکسیمم است اگر و تنها اگر

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}.$$

ماکسیمم مقدار اطلاع متناظر برابر است با

$$H(X) = \log(\sigma\sqrt{2\pi e}). \qquad (۴۵.۵)$$

برهان

اکنون باید چگالی احتمال $p(x)$ را به قسمی تعیین کنیم که

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx$$

ماکسیمم شود، که در آن باید محدودیتهای زیر منظور گردد

$$\int_{-\infty}^{\infty} p(x) dx = 1,$$

و

$$\int_{-\infty}^{\infty} x^2 p(x) dx = \sigma^2,$$

که در آن این فرض شده است σ^2 ثابت است. اکنون تابع زیر را تشکیل می‌دهیم

$$G(x) = -p(x) \log p(x) + \alpha_1 p(x) + \alpha_2 x^2 p(x),$$

و مشتق $G(x)$ نسبت به $p(x)$ را برابر صفر قرار می‌دهیم. این نتیجه می‌دهد

$$-\log p(x) - \log e + \alpha_1 + \alpha_2 x^2 = 0$$

یا پس از تقسیم بر $-\log_2 e$ داریم:

$$\ln p(x) + 1 - \lambda_1 - \lambda_2 . x^2 = 0$$

با

$$\lambda_1 = \frac{\alpha_1}{\log e} \qquad \text{و} \qquad \lambda_2 = \frac{\alpha_2}{\log e}.$$

این جواب زیر را می‌دهد:

$$p(x) = e^{\lambda_1 - 1} e^{\lambda_2 x^2}$$

پارامترهای λ_1 و λ_2 را با قرار دادن $p(x)$ در هر دو شرط حذف می‌کنیم:

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_{-\infty}^{\infty} e^{\lambda_1-1} e^{\lambda_2 x^2}\,dx = 1.$$

که نتیجه می‌شود

$$e^{\lambda_1-1} = \sqrt{-\frac{\lambda_2}{\pi}}.$$

به علاوه،

$$\int_{-\infty}^{\infty} x^2 p(x)dx = \int_{-\infty}^{\infty} x^2 e^{\lambda_2 x^2}\sqrt{-\frac{\lambda_2}{\pi}}\,dx = \sigma^2,$$

که از آن نتیجه می‌شود

$$\lambda_2 = -\frac{1}{2\sigma^2},$$

و بنابراین

$$e^{\lambda_1-1} = \frac{1}{\sigma\sqrt{2\pi}}.$$

که سرانجام نتیجه می‌شود

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}.$$

مقدار اطلاع متناظر برابر است با

$$H(X) = -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}\log\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}dx$$

$$= \log\sigma\sqrt{2\pi}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}dx$$

$$+\int_{-\infty}^{\infty}\frac{x^2\log e}{2\sigma^2}\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}dx$$

$$= \log\sigma\sqrt{2\pi} + \frac{\log e}{2\sigma^2}\mathrm{var}(\mathbf{x}) = \log\sigma\sqrt{2\pi} + \frac{1}{2}\log e$$

$$= \log\sigma\sqrt{2\pi e}.$$

■

از این رو معلوم می‌شود که توزیع نرمال برای یک توان ثابت σ^2 مقدار ماکسیمم

اطلاع را می‌دهد. $H(X)$ متناسب با لگاریتم انحراف معیار σ است. این یک نتیجهٔ مهم است چون توان و توزیع نرمال هر دو غالباً در کاربردهای تکنیکی به کار برده می‌شوند.

۴.۵ اندازه‌های اطلاع و منابع باحافظه

دربارهٔ حالت گسسته علاوه بر اندازهٔ اطلاع حاشیه‌ای اندازه‌های اطلاع شرطی، توأم و متقابل را تعریف کردیم. این کار را برای حالت پیوسته نیز می‌توان انجام داد. اندازه‌های اطلاع شرطی، توأم و متقابل به طور مختصر برای حالت پیوسته در زیر معرّفی شده‌اند. توجّه کنید که رابطهٔ بین اندازه‌های گوناگون روی‌هم‌رفته با رابطه‌های مربوط به حالت گسسته سازگارند.

در حالتی که دو متغیّر تصادفی **x** و **y** با چگالی احتمال توأم $p(x,y)$ وجود دارند ***مقدار اطلاع توأم*** به صورت زیر تعریف می‌شود

$$H(X,Y)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(x,y)\,dx\,dy. \qquad (۴۶.۵)$$

مقدار اطلاع شرطی را می‌توان به صورت زیر تعریف کرد. چگالی احتمال توأم برای دو متغیّر تصادفی **x** و **y** را می‌توان به صورت زیر نوشت

$$p(x,y)=p(x)\,.\,q(y|x)=q(y)\,.\,p(x|y). \qquad (۴۷.۵)$$

اکنون ***مقدار اطلاع شرطی*** x به شرط y به صورت زیر تعریف می‌شود

$$H(X|Y)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log p(x|y)\,dx\,dy. \qquad (۴۸.۵)$$

به طور مشابه

$$H(Y|X)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log q(y|x)\,dx\,dy, \qquad (۴۹.۵)$$

از این رو این تعاریف کاملاً با تعاریف داده شدهٔ قبلی برای حالت گسسته سازگارند. درست مانند حالت گسسته می‌توان نتیجه گرفت که

$$H(X|Y)\le H(X), \qquad (۵۰.۵)$$

$$H(Y|X)\le H(Y), \qquad (۵۱.۵)$$

اگر **x** و **y** به طور آماری مستقل باشند تساوی برقرار می‌باشد. همچنین داریم

$$H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y). \qquad (۵۲.۵)$$

اثباتها را در این جا ارائه نخواهیم داد زیرا اثبات آنها همانند اثباتهای حالت گسسته انجام می‌شوند. براساس مطالب گفته شدهٔ قبلی همچنین داریم

$$H(X,Y)\le H(X)+H(Y), \qquad (۵۳.۵)$$

تساوی در حالتی برقرار است که x و y به طور آماری مستقل باشند. با وجود این نمی‌توانیم بگوییم که $H(Y|X)\ge ۰$. اگر مقدار اطلاع متقابل را به صورت زیر تعریف کنیم

$$I(X;Y)=H(X)+H(Y)-H(X,Y), \qquad (۵۴.۵)$$

در این صورت با کمک تعاریف داده شدهٔ قبلی نتیجه می‌شود که $I(X;Y)$ برابر است با

$$I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log\frac{p(x,y)}{p(x)\cdot q(y)}dx\,dy. \qquad (۵۵.۵)$$

مثال ۱.۵

فرض کنید تابع چگالی احتمال توأم $p(x,y)$ به صورت زیر باشد:

$$p(x,y)=\frac{۱}{۴} \quad , \quad ۰\le x\le ۲ \quad و \quad ۰\le y\le ۴-۲x$$

$$=۰ \quad , \quad \text{در سایر نقاط}$$

برای توابع چگالی احتمال حاشیه‌ای نتیجه می‌شود که

$$p(x)=\int_{۰}^{۴-۲x}p(x,y)dy=\int_{۰}^{۴-۲x}\frac{۱}{۴}dy=\frac{۱}{۴}y\Big|_{۰}^{۴-۲x}=۱-\frac{۱}{۲}x \quad , \quad ۰\le x\le ۲,$$

و

$$q(y)=\int_{۰}^{۲-\frac{۱}{۲}y}p(x,y)dx=\int_{۰}^{۲-\frac{۱}{۲}y}\frac{۱}{۴}dx=\frac{۱}{۴}x\Big|_{۰}^{۲-\frac{۱}{۲}y}=۱-\frac{۱}{۲}x \quad , \quad ۰\le y\le ۴.$$

اکنون می‌توان مقدار اطلاع را مستقیماً محاسبه کرد، به طور کلّی با در نظر گرفتن

$$\int x^n \ln x\,dx=\frac{x^{n+۱}}{n+۱}\left\{\ln x-\frac{۱}{n+۱}\right\}.$$

$H(X)$ را به دست می‌آوریم

$$H(X)=-\int_{۰}^{۲}(۱-\frac{۱}{۲}x)\log(۱-\frac{۱}{۲}x)dx$$

$$= ۴\int_{۱}^{۰} z \log z \, dz = ۲\log e \int_{۱}^{۰} z \ln z \, dz$$

$$= ۲\log e \left[\frac{۱}{۲} z^{۲} \ln z - \frac{۱}{۴} z^{۲}\right]_{۱}^{۰} = \log \sqrt{e} \approx ۰٫۷۲.$$

به طور مشابه می‌توان ثابت کرد که

$$H(Y) = -\int_{۰}^{۴} (\frac{۱}{۲} - \frac{۱}{۸} y) \log(\frac{۱}{۲} - \frac{۱}{۸} y)\, dy = \log ۲\sqrt{e} \approx ۱٫۷۲$$

مقادیر اطلاع شرطی را می‌توان با تعیین کردن توابع چگالی احتمال شرطی محاسبه کرد.

$$p(x|y) = \frac{p(x,y)}{q(y)} = \frac{\frac{۱}{۴}}{\frac{۱}{۲} - \frac{۱}{۸} y} = \frac{۲}{۴ - y} \qquad , \qquad ۰ \le x \le ۲ - \frac{۱}{۲} y,$$

و

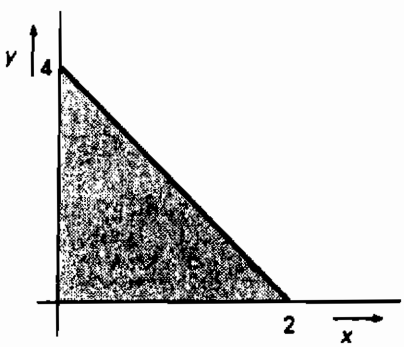
$$p(y|x) = \frac{p(x,y)}{p(x)} = \frac{\frac{۱}{۴}}{۱ - \frac{۱}{۲} x} = \frac{۱}{۴ - ۲x} \qquad , \qquad ۰ \le y \le ۴ - ۲x.$$

با جایگزینی در اندازهٔ اطلاع شرطی نتیجه می‌شود که

$$H(X|Y) = \log \frac{۲}{\sqrt{e}} \approx ۰٫۲۸,$$

و

$$H(Y|X) = \log \frac{۴}{\sqrt{e}} \approx ۱٫۲۸.$$



شکل ۱۲.۵- سطح تعریف $p(x,y)$

این نتایج را می‌توان غیرمستقیم از طریق $H(X,Y)$ به دست آورد.

$$H(X,Y) = -\int_{0}^{2}\int_{0}^{4-2x} p(x,y)\log p(x,y)$$

$$= -\int_{0}^{2}\int_{0}^{4-2x} \frac{1}{4}\log\frac{1}{4}\,dy\,dx$$

$$= \frac{1}{2}\int_{0}^{2}(4-2x)\,dx = \frac{1}{2}[4x - x^2]_0^2 = 2.$$

از $H(X,Y)$ و $H(Y)$ نتیجه می‌شود که

$$H(X|Y) = H(X,Y) - H(Y) = 2 - \log 2\sqrt{e} = \log\frac{2}{\sqrt{e}},$$

و

$$H(Y|X) = H(X,Y) - H(X) = 2 - \log\sqrt{e} = \log\frac{4}{\sqrt{e}}. \qquad \blacktriangle$$

اندازه‌های اطلاع پیوسته توأم و شرطی در تشریح منابع اطلاع پیوسته باحافظه نقشی را بازی می‌کنند.

نمونه‌های حاصل از یک سیگنال تولید شده توسط چنین منبعی عموماً وابسته‌اند. برای منبع اطلاع گسسته این وابستگی با احتمالهای انتقال با نمادها یا حالتهای مختلف در یک زنجیر مارکوف بیان شده است. در حالت اطلاع پیوسته این کار را می‌توان برحسب توابع چگالی احتمال شرطی یا ماتریس خودهمبستگی $\mathbf{R}_{xx}$ انجام داد.

فرض کنید برطبق قضیهٔ نمونه‌گیری $N = 2wT$ نمونه $x(t_1)$، ... ، $x(t_N)$ داریم. به علّت اثر حافظه منبع این نمونه‌ها به طور آماری مستقل نیستند. به موجب این امر باید برای محاسبهٔ مقدار اطلاع توابع چگالی احتمال توأم را به کار برد. این نتیجه می‌دهد

$$H(\tilde{X}) = -\int_{-\infty}^{\infty} p(\tilde{x})\log p(\tilde{x})\,d\tilde{x}$$

$$= -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(x_1,\ldots,x_N)\log p(x_1,\ldots,x_N)\,dx_1\ldots dx_N \qquad (۵۶.۵)$$

که در آن $p(\tilde{x}) = p(x_1,\ldots,x_N) = p(x(t_1),\ldots,x(t_N))$.

مقدار تقریبی مقدار اطلاع برای هر نمونه را می‌توان با تقسیم $H(\tilde{X})$ بر $N = 2wT$ به دست آورد.

به طور کلّی نتیجه برابر $H(X)$ نیست، بلکه مقدار اطلاع تنها یک نمونه می‌باشد. **فقط**

برای منابع اطلاع بی‌حافظه نتیجه می‌شود که

$$\frac{H(\widetilde{X})}{N}=H(X). \tag{۵۷.۵}$$

در این حالت

$$p(x_1,x_2,\ldots,x_N)=\prod_{i=1}^{N}p(x_i)$$

و از این رو

$$\begin{aligned}H(\widetilde{X})&=-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,x_2,\ldots,x_N)\log p(x_1,x_2,\ldots,x_N)\,dx_1\,dx_2\ldots dx_N\\&=\sum_{i=1}^{N}\left\{-\int_{-\infty}^{\infty}p(x_i)\log p(x_i)dx_i\right\}=N\,H(X).\end{aligned}$$

به طور واضح، کمیّت $\frac{H(\widetilde{X})}{N}$ متناظر با $H_N(U)$ در بخش (۲.۳) می‌باشد.

در مقایسه با معادلهٔ (۴۸.۵) می‌توان مقدار اطلاع شرطی یک نمونه به شرط نمونهٔ قبلی را به صورت زیر تعریف کرد

$$H(X_2|X_1)=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x_1,x_2)\log p(x_2|x_1)\,dx_1dx_2. \tag{۵۸.۵}$$

اکنون این مقدار اطلاع شرطی نیز دارای این ویژگی است که

$$H(X_2|X_1)\le H(X_2), \tag{۵۹.۵}$$

و از این رو

$$\begin{aligned}H(X_1,X_2)&=H(X_1)+H(X_2|X_1)\\&=H(X_2)+H(X_1|X_2)\le H(X_1)+H(X_2),\end{aligned} \tag{۶۰.۵}$$

برابری برقرار است اگر نمونه‌ها به طور آماری مستقل باشند (منبع بی‌حافظه). بنابراین آگاهی از x_1 موجب کاهش عدم‌حتمیت دربارهٔ x_2 می‌شود. اگر $2NT$ نمونه در نظر بگیریم بنابر بخش (۲.۳) تعریف می‌کنیم

$$F_N(X)=H(X_N|X_{N-1},\ldots,X_1)\ \text{بیت / نماد}, \tag{۶۱.۵}$$

و

$$H_N(X)=\frac{1}{N}H(X_1,\ldots,X_N)=\frac{1}{N}H(\widetilde{X})\ \text{بیت / نماد}, \tag{۶۲.۵}$$

که مقدار اطلاع در هر نمونه است که در آن استقلال نمونه‌ها مورد توجّه قرار گرفته است.

اگر چگالی احتمال گاوسی باشد، عبارات گوناگون بیشتری می‌توان پیدا کرد. برای چگالیهای احتمال دیگر اغلب این کار خیلی مشکل است.

مقدار اطلاع منبع گاوسی مانا به صورت زیر می‌باشد

$$H(\widetilde{X}) = N \log\{\sigma\sqrt{2\pi e}\} \qquad (63.5)$$

اگر سیگنال از N نمونهٔ ناهمبسته تشکیل شده باشد (معادلهٔ (۴۴.۵) را ملاحظه کنید) برای نمونه‌های وابسته با چگالی احتمال N- بعدی $p(\widetilde{x})$ داریم

$$H(\widetilde{X}) = \log\{(2\pi e)^{N/2}|\mathbf{K}_{xx}|^{\frac{1}{2}}\}, \qquad (64.5)$$

که در آن $|\mathbf{K}_{xx}|$ دترمینان ماتریس اتوکوواریانس $\mathbf{K}_{xx}$ است. اگر نمونه‌ها ناهمبسته باشند، در این صورت اگر $\mathbf{K}_{xx}$ یک ماتریس قطری با عناصر قطری σ^2 باشد، حالت یادآوری شدهٔ قبل مستقیماً به دست می‌آید. در مقایسه با حالت یک-بعدی می‌توان نشان داد که اگر منبع گاوسی باشد از یک منبع اطلاع با ماتریس اتوکوواریانس داده شده ماکسیمم مقدار اطلاع به دست می‌آید.

مثال ۲.۵

فرض کنید $N=2$ و تابع چگالی احتمال گاوسی دو-بعدی با $\mu_1=\mu_2=0$، $\sigma_1=\sigma_2=1$ و $\rho=\frac{1}{2}$ باشد. با کمک معادلهٔ (۱۹.۵) به دست می‌آوریم

$$p(x_1, x_2) = \frac{1}{\pi\sqrt{3}} e^{-\frac{2}{3}(x_1^2 - x_1x_2 - x_2^2)},$$

و چون

$$p(x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2}$$

تابع چگالی احتمال شرطی برابر می‌شود با

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)} = \frac{1}{\frac{1}{2}\sqrt{6\pi}} e^{-\frac{2}{3}(x_2-\frac{1}{2}x_1)^2}.$$

این مجدداً یک توزیع گاوسی است؛ به طوری که $\mu=\frac{1}{2}x_1$ و $\sigma=\frac{1}{2}\sqrt{3}$، و از این رو

$$H(X_2|X_1) = \log\sigma\sqrt{2\pi e} = \log\frac{1}{2}\sqrt{6\pi e} \approx 1{,}84,$$

$H(\widetilde{X})$ برابر می‌شود با

$$H(\widetilde{X}) = H(X_1, X_2) = H(X_1) + H(X_2 | X_1) = \log\sqrt{2\pi e} + \log\frac{1}{2}\sqrt{6\pi e}$$
$$= \log(\pi e\sqrt{3}).$$

اگر معادلهٔ (۶۴.۵) را به کار بریم همین نتیجه مستقیماً به دست می‌آید. چون $\sigma_1 = \sigma_2 = 1$ و $\rho = \frac{1}{2}$ با معادلهٔ (۲۷.۵) نتیجه می‌شود که $\operatorname{cov}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2}$. با به کار بردن معادلات (۳۴.۵) و (۳۵.۵) ماتریس اتوکوواریانس به صورت زیر به دست می‌آید

$$K_{xx} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

بنابراین،

$$H(\widetilde{X}) = \log\{(2\pi e)^{N/2} |\mathbf{K}_{xx}|^{\frac{1}{2}}\} = \log\left\{2\pi e \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}^{1/2}\right\}$$
$$= \log(\pi e\sqrt{3}) \approx 3{,}89.$$

▲

در پایان این بخش تذکّراتی نسبت به تبدیل متغیّرهای تصادفی ارائه خواهد شد. در حالت تبدیل متغیّر تصادفی $\mathbf{x}$ به متغیّر تصادفی $\mathbf{y}$ با ضابطهٔ $\mathbf{y} = f(\mathbf{x})$ تابع چگالی احتمال و همچنین مقدار اطلاع تغییر خواهد کرد. در حالت کلّی تابع چگالی احتمال N- بعدی موردی است که

$$q(y_1, \dots, y_N) = p(x_1, \dots, x_N) \left| J\left(\frac{x_1 \dots x_N}{y_1 \dots y_N}\right) \right| \qquad (۶۵.۵)$$

بنابراین

$$\int_{x_1}\int_{x_2}\dots\int_{x_N} p(x_1, \dots, x_N)\, dx_1\, dx_2 \dots dx_N = 1,$$

و

$$J\left(\frac{x_1 \dots x_N}{y_1 \dots y_N}\right)$$

را ژاکوبی x_1، x_2، ...، x_N نسبت به y_1، y_2، ...، y_N می‌نامیم.

برای $N = 3$ ژاکوبی به صورت زیر تعریف می‌شود

$$J\left(\frac{x_1,x_2,x_3}{y_1,y_2,y_3}\right)=\begin{bmatrix}\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3}\\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3}\\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3}\end{bmatrix} \tag{۶۶.۵}$$

توجّه کنید که

$$\left|J\left(\frac{x_1\ldots x_N}{y_1\ldots y_N}\right)\right|=\frac{1}{\left|J\left(\frac{y_1\ldots y_N}{x_1\ldots x_N}\right)\right|}. \tag{۶۷.۵}$$

برای مقدار اطلاع $H(\tilde{Y})$ به دست می‌آوریم

$$H(\tilde{Y})=-\int_{y_1}\int_{y_2}\ldots\int_{y_N} p(x_1,\ldots,x_N)\left|J\left(\frac{x_1\ldots x_N}{y_1\ldots y_N}\right)\right| \times\log\left\{p(x_1,\ldots,x_N)\left|J\left(\frac{x_1\ldots x_N}{y_1\ldots y_N}\right)\right|dy_1\ldots dy_N\right\}. \tag{۶۸.۵}$$

در حالت یک-بعدی نتیجه می‌شود

$$\begin{aligned}H(Y)&=-\int_y q(y)\log q(y)\,dy\\ &=-\int_x p(x)J(x|y)\log\{p(x)J(x|y)\}J(x|y)\,dx\\ &=H(X)-E_x\{\log J(x|y)\}.\end{aligned} \tag{۶۹.۵}$$

از این رو مقدار اطلاع $H(Y)$ به استثنای جملهٔ ثابت با $H(X)$ برابر است.

توجّه خاصّی باید به مواردی معطوف داشت که در آن باید تابع چگالی احتمال یک متغیّر تصادفی را تعیین کرد؛ برای مثال هنگامی که متغیّر تصادفی مجموع دو متغیّر تصادفی دیگر است.

فرض کنید $\mathbf{x}$ و $\mathbf{y}$ متغیّرهای تصادفی باشند و $\mathbf{z}$ متغیّر تصادفی است به طوری که $\mathbf{z}=\mathbf{x}+\mathbf{y}$. توابع چگالی احتمال نسبت به $\mathbf{x}$ و $\mathbf{y}$ به ترتیب با $p(x)$ و $p(y)$ داده شده‌اند. برای تابع توزیع تجمّعی داریم

$$F(z)=P(x+y\le z)=\int_{-\infty}^{\infty}\int_{-\infty}^{z-x}p(x,y)\,dx\,dy.$$

بنابراین

$$p(z)=\frac{dF(z)}{dz}=\int_{-\infty}^{\infty}\frac{d}{dz}\left\{\int_{-\infty}^{z-x}p(x,y)\,dy\right\}dx.$$

چون به طور کلّی داریم

$$\frac{d}{du}\int_{-\infty}^{u}f(r)\,dr=f(u)-f(-\infty), \tag{۷۰.۵}$$

و x و y مستقلند، در این حالت داریم

$$p(z)=\int_{-\infty}^{\infty}p(x,z-x)\,dx=\int_{-\infty}^{\infty}p(x)p_y(z-x)\,dx. \tag{۷۱.۵}$$

مثال ۳.۵

فرض کنید:

$$p(x)=\frac{۱}{۲} \quad , \quad ۰\le x\le ۲$$

$$p(y)=\frac{۱}{۲} \quad , \quad ۰\le y\le ۲$$

و $\mathbf{z}=\mathbf{x}+\mathbf{y}$.

چون $۰\le y\le ۲$، $۰\le z-x\le ۲$ و از این رو $z-۲\le x\le z$، برای $p(z)$ دو حالت تشخیص داده می‌شود.

(الف) $z\le ۲$ چون هر دوی x و y بزرگتر از صفرند، بنابراین $z\ge ۰$. از این رو $۰\le z\le ۲$.

برای x نتیجه می‌شود که $۰\le x\le z$ و $z-۲\le x\le z$. شکل (۱۳.۵- الف) را ملاحظه کنید. ترکیب این دو به $۰\le x\le z$ منجر می‌شود. برای $p(z)$ به دست می‌آوریم

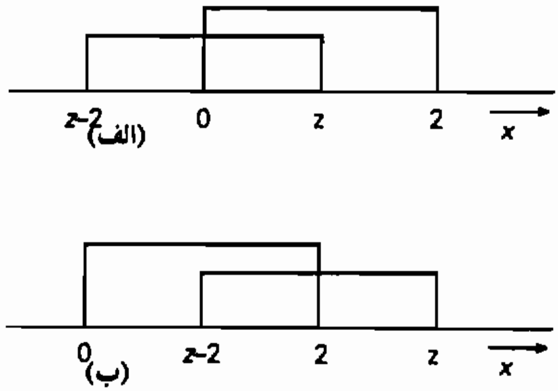
$$p(z)=\int_{۰}^{z}\frac{۱}{۲}\times\frac{۱}{۲}\,dx=\frac{۱}{۴}x\Big|_{۰}^{z}=\frac{۱}{۴}z.$$

(ب) $z\ge ۲$ چون ماکسیمم مقدار x و y برابر ۲ است، داریم: $۲\le z\le ۴$.

اکنون $z-۲\le x\le ۲$ (شکل ۱۳.۵- ب) را ببینید) و از این رو

$$p(z)=\int_{z-۲}^{۲}\frac{۱}{۲}\times\frac{۱}{۲}\,dx=\frac{۱}{۴}x\Big|_{z-۲}^{۲}=۱-\frac{۱}{۴}z.$$

برای $H(Z)$ به دست می‌آوریم



شکل ۱۳.۵

$$H(Z)=-\int_{0}^{2}\frac{1}{4}z\log\frac{1}{4}z\,dz-\int_{2}^{4}(1-\frac{1}{4}z)\log(1-\frac{1}{4}z)\,dz$$

$$=-4\int_{0}^{\frac{1}{2}}u\log u\,du+4\int_{\frac{1}{2}}^{0}u\log u\,du$$

$$=-8\int_{0}^{\frac{1}{2}}u\log u\,du=-\frac{8}{\ln 2}u^{2}(\frac{1}{2}\ln u-\frac{1}{4})\Big|_{0}^{\frac{1}{2}}$$

$$=1+\frac{1}{2}\log e.$$ ▲

۵.۵ توان اطلاع

در تعریف (۱.۵) از بخش (۳.۵) مقدار اطلاع برای یک منبع بی‌حافظه پیوسته داده شد. همچنان که در همین بخش دیدیم ماکسیمم مقدار اطلاعی که یک منبع می‌تواند بدهد بستگی به قیود دارد. برای مثال اگر متوسط توان سیگنال (واریانس) تولید شده توسط منبع از قبل تعیین شده باشد نتیجه می‌شود که $\max H(X)=\log[\sigma\sqrt{2\pi e}]$. اگر تابع چگالی احتمال $p(x)$ منبع اطلاع پیوسته گاوسی باشد با واریانس σ^2، این مقدار ماکسیمم به دست می‌آید. همان‌طور که قبلاً تذکر داده شد، منبع اطلاع گاوسی برای توصیف سیگنالهای تصادفی پیوسته اهمیت زیادی دارد. این الگوی همراه با منبع اطلاع گسستهٔ دودویی در بین الگوها بیشترین کاربرد را در نظریهٔ اطلاع دارند. از دید این حقیقت که مقدار ماکسیمم

اطلاع بستگی به قیود داده شده دارد، مفهوم حشو برای منابع پیوسته را باید با دقّت بیشتری بررسی کرد. می‌توان تنها از حشو صحبت کرد اگر قیود فرض شده مورد توجّه قرار گیرند. در حالت توانهای محدود، حشو را می‌توان به صورت زیر تعریف کرد

$$red = ۱ - \frac{H(X)}{\log[\sigma\sqrt{۲\pi e}]}. \qquad (۷۲.۵)$$

منابع زیادی وجود دارند که توان کراندار دارند ولی گاوسی نیستند. اکنون می‌توان به دست آورد که یک منبع گاوسی در مقایسه با یک منبع دلخواه در حالتی که هر دو دارای یک مقدار اطّلاعند باید چه مقدار توان داشته باشد. این موضوع به مفهوم ***توان اطلاع*** منجر می‌شود.

توان اطلاع P_H یک سیگنال تصادفی $\mathbf{x}(t)$ تولیده شده توسط یک منبع اطلاع برابر با توان یک سیگنال گاوسی است که دارای همان مقدار اطلاعی است که سیگنال تصادفی دارد.

فرض کنید $H(X)$ مقدار اطلاع منسوب به سیگنال تصادفی $\mathbf{x}(t)$ باشد. بنابر تعریف توان اطلاع مقداری از $\sigma^۲$ را پیدا می‌کنیم که برای آن مقدار اطلاع منبع گاوسی با این مقدار $\sigma^۲$ برابر $H(X)$ می‌شود. در این صورت توان اطلاع P_H برابر مقدار $\sigma^۲$ است. از

$$H(X) = \log\sigma\sqrt{۲\pi e} = \log\sqrt{۲\pi e P_H} \qquad (۷۳.۵)$$

نتیجه می‌شود که

$$P_H = \frac{۱}{۲\pi e}۲^{۲H(X)}, \qquad (۷۴.۵)$$

که بیان ریاضی توان اطلاع در حالت سیگنالهای دودویی با مقدار اطلاع $H(X)$ است.

چون سیگنال گاوسی به ماکسیمم مقدار اطلاع برای توان داده شده منجر می‌شود، توان اطلاع یک سیگنال دلخواه همواره از حالت گاوسی کوچکتر است.

مثال ۴.۵

فرض کنید تابع چگالی احتمال $p(x)$ به صورت زیر باشد

$$p(x) = \frac{x^۳}{۲۵۰۰} \quad , \quad ۰ \le x \le ۱۰$$

$$= ۰ \quad , \quad \text{سایر نقاط}$$

مقدار اطلاع $H(X)$ برابر است با

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx$$
$$= -\int_{۰}^{۱۰}\frac{x^{۳}}{۲۵۰۰}\log\frac{x^{۳}}{۲۵۰۰}dx$$
$$= -\frac{۳}{۲۵۰۰}\int_{۰}^{۱۰}x^{۳}\log x\,dx + \frac{\log ۲۵۰۰}{۲۵۰۰}\int_{۰}^{۱۰}x^{۳}\,dx$$
$$= -\frac{۳\log e}{۲۵۰۰}\int_{۰}^{۱۰}x^{۳}\ln x\,dx + \frac{\log ۲۵۰۰}{۲۵۰۰}\int_{۰}^{۱۰}x^{۳}\,dx$$
$$= -\frac{۳\log e}{۲۵۰۰}\left[\frac{x^{۴}}{۴}(\ln x - \frac{۱}{۴})\right]_{۰}^{۱۰} + \frac{\log ۲۵۰۰}{۲۵۰۰}\left[\frac{۱}{۴}x^{۴}\right]_{۰}^{۱۰}$$
$$= \log\frac{۵}{۲}\sqrt[۴]{e^{۳}} \approx ۲{,}۴۰.$$

برای میانگین و واریانس به ترتیب نتیجه می‌شود که

$$\mu = \int_{۰}^{۱۰}\frac{x^{۴}}{۲۵۰۰}dx = \left[\frac{x^{۵}}{۱۲۵۰۰}\right]_{۰}^{۱۰} = ۸,$$

$$\sigma^{۲} = \int_{۰}^{۱۰}(x-۸)^{۲}\frac{x^{۳}}{۲۵۰۰}dx = \frac{۱}{۲۵۰۰}\int_{۰}^{۱۰}x^{۵}\,dx - \frac{۶۴}{۲۵۰۰}\int_{۰}^{۱۰}x^{۳}\,dx$$
$$= \left[\frac{x^{۶}}{۱۵۰۰۰} - \frac{۱۶x^{۴}}{۲۵۰۰}\right]_{۰}^{۱۰} = \frac{۸}{۳}.$$

با فرض $\sigma^{۲} = \frac{۸}{۳}$، مقدار ماکسیمم اطلاع برابر است با

$$\log\sigma\sqrt{۲\pi e} = \log\sqrt{۲\pi e\frac{۸}{۳}} \approx ۲{,}۷۵.$$

به عنوان یک نتیجه حشو منبع اطلاع برابر می‌شود با

$$red = ۱ - \frac{H(X)}{\log\sigma\sqrt{۲\pi e}} = ۱ - \frac{۲{,}۴۰}{۲{,}۷۵} \approx ۰{,}۱۳.$$

برای توان اطلاع به دست می‌آوریم

$$P_H = \frac{۱}{۲\pi e}۲^{۲H(X)} = \frac{۱}{۲\pi e}\left\{\frac{۵}{۲}\sqrt[۴]{e^{۳}}\right\}^{۲} = \frac{۲۵\sqrt{e}}{۸\pi} = ۱{,}۶۴.$$

واضح است که توان اطلاع واقعاً از توان واقعی که برابر است بـا $P = \sigma^{۲} = \frac{۸}{۳} \approx ۲{,}۶۶$

کوچکتر است. ▲

مفهوم توان اطلاع مهم است وقتی یک کانال ارتباطی پیوسته را توصیف می‌کنیم که در آن نوفه با چگالی احتمال غیرگاوسی حضور دارد.

در حالت کلّی کـه بـا مجمـوع $\mathbf{z}(t)$ از دو سـیگنال $\mathbf{x}(t)$ و $\mathbf{y}(t)$، $\mathbf{z}(t)=\mathbf{x}(t)+\mathbf{y}(t)$ سروکار داریم، می‌توان نشان داد که نابرابری زیر برقرار است:

$$P_{H_x}+P_{H_y}\le P_{H_z}\le P_x+P_y, \qquad (۷۵.۵)$$

که در آن P_{H_x}، P_{H_y} و P_{H_z} به ترتیب توان اطلاع سه سیگنال $\mathbf{x}(t)$، $\mathbf{y}(t)$ و $\mathbf{z}(t)$ اسـت. P_x و P_y به ترتیب توانهای $\mathbf{x}(t)$ و $\mathbf{y}(t)$ هستند. اگر هر دو سـیگنال گاوسـی باشـند علامـت برابری در بالا برقرار است زیرا در این حالت توان اطلاع و توان واقعی یکسان هستند.

واقعاً آنچه نابرابری بیان می‌کند این است که اگر دو سیگنال بـه هـم اضافـه شـوند مقدار اطلاع افزایش می‌یابد یا حدّاقل با روی هم ریختن کاهش نمی‌یابد. این مطلب با بیان توان (اطلاع) برحسب اندازه‌های اطلاع نتیجه می‌شود. ایـن را کـه مقـدار اطـلاع افزایـش خواهد یافت یا حدّاقل در حالت افزودن بر یکدیگر کـاهش نخواهـد یـافت می‌تـوان بـه صورت زیر روشن کرد. اگر دو سیگنال مستقل $\mathbf{x}(t)$ و $\mathbf{y}(t)$ را به هم اضافه کنیم، ممکـن است انتظار داشته باشیم که $\mathbf{z}(t)$، مجموع آنها، میل خواهد داشت که سیگنال نوفه-تصادفی بشود. چون نوفهٔ تصادفی با توزیع گاوسی مشخّص شده است این بدین معناست کـه تـابع چگالی احتمال $\mathbf{z}(t)$ به یک توزیع گاوسی می‌گراید. این حقیقت که توزیـع گاوسـی بـه مقدار اطلاع ماکسیمم منجر می‌شود دارای نتیجه‌ای است که در حالت افزودن به یکدیگـر ممکن است افزایش مقدار اطلاع را انتظار داشته باشیم.

مثال ۵.۵

فرض کنید دو منبع اطلاع پیوسته با توابع چگالی احتمال زیر داشته باشیم

$$p(x)=\frac{۱}{۲} \quad , \quad ۰\le x\le ۲$$
$$=۰ \quad , \quad \text{سایر نقاط}$$

$$p(y)=\frac{۱}{۲} \quad , \quad ۰\le x\le ۲$$
$$=۰ \quad , \quad \text{سایر نقاط}$$

این متناظر است با $H(X)=H(Y)=۱$ و از این رو

$$P_{H_x} = P_{H_y} = \frac{۱}{۲\pi e} ۲^{۲H(X)} = \frac{۲}{\pi e}.$$

توان واقعی برابر است با $P_x = P_y = \frac{۱}{۳}$، همچنان که می‌توان از $\sigma_x^۲$ محاسبه کرد
اگر x و y مستقل باشند تابع چگالی احتمال نسبت به $z = x + y$ به صورت زیر داده می‌شود (مثال (۳.۵) را ببینید)

$$\begin{aligned} p(z) &= \frac{۱}{۴} z \qquad , \quad ۰ \le z \le ۲, \\ &= ۱ - \frac{۱}{۴} z \quad , \quad ۲ \le z \le ۴, \\ &= ۰ \qquad , \quad \text{سایر نقاط}. \end{aligned}$$

اگر $H(Z)$ را محاسبه کنیم به دست می‌آوریم

$$H(Z) = \log ۲\sqrt{e},$$

و

$$P_z = \sigma_z^۲ = \frac{۲}{۳}.$$

برای اطلاع چنین داریم که

$$P_{H_z} = \frac{۱}{۲\pi e} ۲^{۲\log ۲\sqrt{e}} = \frac{۲}{\pi} \approx ۰{,}۶۴.$$

اگر این را با نابرابری فرمول (۷۵.۵) مقایسه کنیم داریم

$$P_{H_x} + P_{H_y} = \frac{۴}{\pi e} = ۰{,}۴۷ \le P_{H_z} = \frac{۲}{\pi} = ۰{,}۶۴ \le P_x + P_y = \frac{۲}{۳} = ۰{,}۶۷. \qquad \blacktriangle$$

۶.۵ تمرینها

۱.۵ (الف) متغیّر تصادفی x دامنهٔ نوسان سیگنال $x(t)$ را که بین ۳− و ۳+ ولت محدود شده است نشان می‌دهد. x دارای توزیع احتمال یکنواخت بین این حدود است. مقدار اطلاع $H(X)$ را تعیین کنید.

(ب) مقدار اطلاع $H(X)$ را اگر x بین ۵− و ۵+ به طور یکنواخت توزیع شده باشد بیابید.

(پ) اختلاف بین جوابهای یافته شده در (الف) و (ب) را تفسیر کنید.

۲.۵ یک نمونهٔ x از یک سیگنال تصادفی به طور یکنواخت بین ۱+ و ۷+ ولت توزیع

شده است.

(الف) مقدار اطلاع $H(X)$ را تعیین کنید. اگر این نتیجه را با نتیجهٔ تمرین (۵.۱-الف) مقایسه کنید چه نتیجه‌ای می‌توانید بگیرید؟

(ب) $E(\mathbf{x})$ و $\mathrm{var}(\mathbf{x})$ را بیابید.

۳.۵ متغیّر تصادفی پیوسته $\mathbf{x}$ با چگالی احتمال $p(x)$ داده شده است؛ برای متغیّر تصادفی $\mathbf{y}$ داریم:

$$\mathbf{y} = \mathbf{x} + \alpha .$$

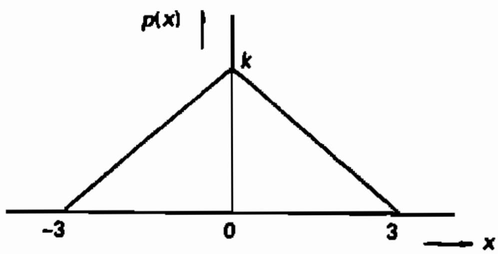
ثابت کنید $H(X) = H(Y)$.

۴.۵ متغیّر تصادفی $\mathbf{x}$ دارای چگالی احتمال به صورت داده شده در شکل (۵.۱۴) است.

(الف) مقدار k را تعیین کنید؛

(ب) $H(X)$ را تعیین کنید؛

(پ) مقدار به‌دست آمده برای $H(X)$ را با جواب تمرین (۵.۱- الف) مقایسه کنید و برای هر اختلافی تفسیری بیان کنید.



شکل ۵.۱۴- چگالی احتمال $p(x)$ تمرین (۵.۴)

۵.۵ یک نمونهٔ $\mathbf{x}$ از سیگنال تصادفی $\mathbf{x}(t)$ به طور یکنواخت بین $-A$ و $+A$ با $A > 0$ توزیع شده است.

(الف) $H(X)$ را به عنوان تابعی از A رسم کنید؛

(ب) $\mathrm{var}(\mathbf{x})$ را به عنوان تابعی از A تعیین نموده و آن را رسم کنید؛

(پ) دو نمودار چه اشتراکی دارند؟

۶.۵ متغیّر تصادفی $\mathbf{x}$ دارای توزیع نمایی منفی است:

$$p(x) = \frac{1}{\lambda}\exp(-\frac{x}{\lambda}) \quad , \quad x \geq 0$$

$$= 0 \quad , \quad x < 0$$

(الف) مقدار اطلاع $H(X)$ را محاسبه کنید؛

(ب) متغیّر تصادفی در نظر بگیرید که نمی‌تواند مقادیر منفی اختیار کند و میـانگین آن λ است؛ نشان دهید که این متغیّر وقتی چگالی احتمالش به صـورت توزیـع داده شده در بالا باشد دارای ماکسیمم مقدار اطلاع است.

۷.۵ یک نمونه x از یک سیگنال $x(t)$ دارای چگالی احتمال زیر است:

$$p(x)=\frac{1}{2\lambda}\exp(-\frac{|x|}{\lambda}).$$

برای یک نمونهٔ y از سیگنال خروجی داریم:

$$\mathbf{y}=|\mathbf{x}|.$$

(الف) $H(X)$ را بیابید؛

(ب) $q(y)$ را تعیین کنید و از این $H(Y)$ را بیابید؛

(پ) ثابت کنید $H(Y|X)=0$.

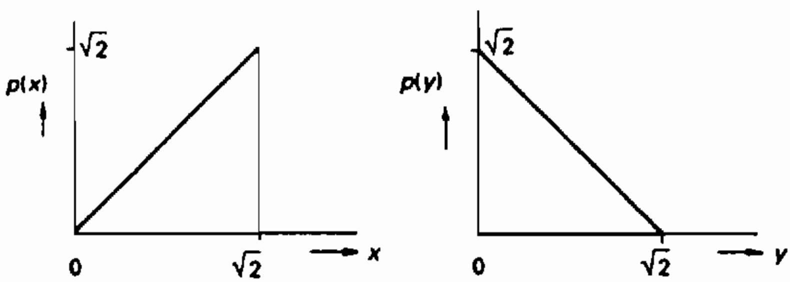
(ت) از جوابهای قبل $H(X|Y)$ را بیابید. آیا می‌توانید توضیحی در مورد مقدار به دسـت آمده ارائه نمایید؟

۸.۵ متغیّرهای مستقل آماری x و y، کـه سـیگنالهای $x(t)$ و $y(t)$ را نشـان می‌دهنـد، دارای چگالیهای احتمال نشان داده شده در شکل (۱۵.۵) می‌باشند.

سیگنالهای $x(t)$ و $y(t)$ را به یک تمیزدهنده (شکل (۱۶.۵) را ببینید) اضافـه می‌کنیم که در آن برای نمونه z سیگنال خروجی داریم

$$\mathbf{z}=\mathbf{x} \quad , \quad \mathbf{x}\geq\frac{1}{2}\sqrt{2},$$

$$\mathbf{z}=\mathbf{y} \quad , \quad \mathbf{x}<\frac{1}{2}\sqrt{2}.$$



شکل ۵.۱۵- چگالی احتمال $p(x)$ و $p(y)$ تمرین (۸.۵)

شکل ۱۶.۵- تمیز دهنده تمرین (۸.۵)

(الف) $H(Z|x < \frac{۱}{۲}\sqrt{۲})$ را تعیین کنید؛

(ب) $H(Z|x \geq \frac{۱}{۲}\sqrt{۲})$ را بیابید؛

(پ) $H(Z)$ را تعیین کنید.

برای بند (ب) باید در نظر داشته باشید که $p(x) \neq p(x|x \geq \frac{۱}{۲}\sqrt{۲})$.

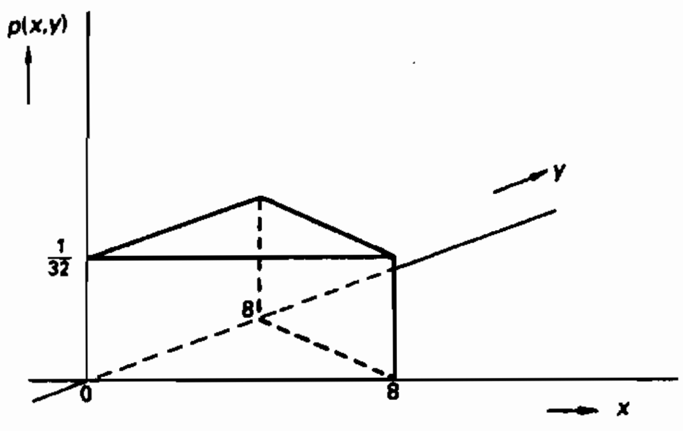
۹.۵ متغیّرهای تصادفی x و y دارای چگالی احتمال توأم رسم شده در شکل (۱۷.۵) می‌باشند.

(الف) $H(X)$ را تعیین کنید؛

(ب) $H(Y|X)$ را بیابید؛

(پ) کمیّت $H(X|Y)$ را از (الف) و (ب) تعیین کنید؛ $E(X|Y)$ را مستقیماً محاسبه نموده و جوابها را با هم مقایسه کنید.

(ت) براساس جوابها نتیجه بگیرید x و y به طور آماری مستقلّند.



شکل ۱۷.۵- چگالی احتمال توأم تمرین (۹.۵)

۱۰.۵ برای متغیّر تصادفی x و y داده شده است که

- $p(x,y)$ چگالی احتمال دو-بعدی است؛
- $E(\mathbf{x}) = E(\mathbf{y}) = 0$؛
- $\mathrm{var}(\mathbf{x}) = \sigma_1^2$ و $\mathrm{var}(\mathbf{y}) = \sigma_2^2$؛
- ρ ضریب همبستگی است.

(الف) $H(Y)$ را به دست آورید؛
(ب) $H(X|Y)$ را تعیین کنید؛
(پ) $H(X,Y)$ را محاسبه کنید؛
(ت) اگر $\rho = 0$ اثر آن بر این کمیتها چیست و چگونه این را توضیح می‌دهید؟

۱۱.۵ یک منبع اطلاع پیوسته سیگنال تصادفی $\mathbf{x}(t)$ را با چگالی احتمال زیر تولید می‌کند:

$$p(x) = e^{-a|x|}$$

(الف) مقدار اطلاع در یک نمونه از این سیگنال را محاسبه کنید؛
(ب) با توان اطلاع چه می‌توان فهمید؟
(پ) توان اطلاع را برای سیگنال بالا محاسبه کنید.

۱۲.۵ یک منبع اطلاع پیوسته سیگنال تصادفی $\mathbf{x}(t)$ را با چگالی احتمال زیر تولید می‌کند:

$$p(x) = ax^2 \qquad , \qquad 0 \le x \le \lambda$$

(الف) مقدار اطلاع $H(X)$ را برای یک نمونه محاسبه کنید و $H(X)$ را به عنوان تابعی از λ $(\lambda > 0)$ رسم کنید.
(ب) برای چه مقدار λ، $H(X) = 0$ است؟ آیا این مقدار λ معنی خاصی دارد؟
(پ) توابع اطلاع این منبع را محاسبه کنید.

۱۳.۵ یک منبع اطلاع پیوسته سیگنال تصادفی $\mathbf{x}(t)$ را با چگالی احتمال زیر تولید می‌کند:

$$p(x) = \begin{cases} \dfrac{1 - \frac{|x|}{a}}{a} , & |x| \le a \\ = 0 , & |x| > a. \end{cases}$$

(الف) مقدار اطلاع را در یک نمونه از این سیگنال محاسبه کنید؛
(ب) توان اطلاع منبع را تعیین کنید؛
(پ) عبارتی برای توان منبع به صورت تابعی از a $(a>0)$ بیابید؛
(ت) $H(X)$ را با مقدار اطلاع $H(Y)$ یک منبع گاوسی، که دارای همان توان منبع تحت بررسی است، مقایسه کنید.

۷.۵ جوابها

۱.۵ (الف) به طور کلّی، برای متغیّر تصادفی $\mathbf{x}$ که محدود بین $-A$ و $+A$ است، در حالی که دارای توزیع یکنواخت بین این دو حدود است، داریم:

$$p(x)=\frac{1}{2A} \qquad , \qquad -A\le x\le A$$

براساس قضیهٔ (۲.۵) مقدار اطلاع مربوط برابر است با

$$H(X)=\log 2A.$$

در این حالت $A=3$، از این رو داریم

$$H(X)=\log 6=2{,}58 \text{ بیت}.$$

(ب) با جایگزینی $A=5$ در معادلهٔ کلّی داده شده نتیجه می‌شود

$$H(X)=\log 10=3{,}32 \text{ بیت}.$$

(پ) می‌توان نتیجه گرفت اگر دامنهٔ متغیّر بزرگتر باشد مقدار اطلاع افزایش خواهد یافت. این با این امر که عدم‌حتمیت مربوط به متغیّر برای دامنه بزرگتر افزایش می‌یابد مطابقت دارد.

۲.۵ (الف) چون با توزیع یکنواخت سروکار داریم نتیجه می‌شود که $p(x)=\frac{1}{6}$ برای $1\le x\le 7$. مقدار اطلاع برابر است با

$$H(X)=-\int_1^7 p(x)\log p(x)dx=-\int_1^7\frac{1}{6}\log\frac{1}{6}dx=\log 6=2{,}58 \text{ بیت}.$$

مقایسه با نتیجهٔ تمرین (۱.۵- الف) نشان می‌دهد که مقدار اطلاع برای توزیع یکنواخت تنها به دامنهٔ متغیّر وابسته است و به موقعیت دامنه بستگی ندارد.

(ب) داریم

$$E(\mathbf{x})=\int_{۱}^{۷} x\, p(x)\,dx=\int_{۱}^{۷}\frac{x}{۶}\,dx=\frac{۱}{۱۲}x^{۲}\Big|_{۱}^{۷}=۴.$$

برای واریانس داریم $\operatorname{var}(\mathbf{x})=E[(\mathbf{x}-E(\mathbf{x}))^{۲}]$.

چون $E(\mathbf{x})=۴$، نتیجه می‌شود که

$$\operatorname{var}(\mathbf{x})=E[(\mathbf{x}-۴)^{۲}]=\int_{۱}^{۷}(x-۴)^{۲}\,p(x)\,dx$$

$$=\int_{۱}^{۷}\frac{(x-۴)^{۲}}{۶}\,dx=\int_{-۳}^{۳}\frac{y^{۲}}{۶}\,dy=\frac{۱}{۱۸}y^{۳}\Big|_{-۳}^{۳}=۳.$$

۳.۵ چون α یک ثابت است، افزایش α به $\mathbf{x}$ عدم‌حتمیت در $\mathbf{y}=\mathbf{x}+\alpha$ را تغییر نخواهد داد. بنابراین مقدار اطلاع نیز بدون تغییر باقی می‌ماند و خواهیم داشت $H(X)=H(Y)$. این مطلب را با این اندیشه که عدم‌حتمیت نسبت به $\mathbf{y}$ در واقع برابر با عدم‌حتمیت نسبت به $\mathbf{x}$ و α، $H(X,\alpha)$، است نیز می‌توان دید.

چون α ثابتی است که مقدار آن از قبل تعیین شده است داریم $H(\alpha)=۰$.

اکنون نتیجه می‌شود که

$$H(Y)=H(X,\alpha)=H(X)+H(\alpha)=H(X).$$

۴.۵ (الف) باید داشته باشیم

$$\int_{-\infty}^{\infty}p(x)\,dx=۱$$

به آسانی دیده می‌شود که k بایستی برابر $\frac{۱}{۳}$ باشد.

(ب) برای این که قادر باشیم $H(X)$ را محاسبه کنیم، ابتدا باید $p(x)$ را پیدا کنیم. می‌توان ثابت کرد که

$$p(x)=\frac{x+۳}{۹}\qquad,\qquad x\le ۰$$

و

$$p(x)=\frac{۳-x}{۹}\qquad,\qquad x>۰$$

برای $H(X)$ نتیجه می‌شود که

$$H(X)=-\int_{-۳}^{۰}\frac{x+۳}{۹}\log\left(\frac{x+۳}{۹}\right)dx-\int_{۰}^{۳}\frac{۳-x}{۹}\log\left(\frac{۳-x}{۹}\right)dx$$

$$= -۱۸\int_{\circ}^{۱/۳} y\log y\,dy.$$

چون به طور کلّی داریم،

$$\int y\ln y\,dy = y^{۲}\left[\frac{۱}{۲}\ln y - \frac{۱}{۴}\right],$$

برای $H(X)$ نتیجه می‌شود که

$$H(X) = -\frac{۱۸}{\ln ۲}\int_{\circ}^{۱/۳} y\ln y\,dy = -\frac{۱۸}{\ln ۲}y^{۲}\left(\frac{۱}{۲}\ln y - \frac{۱}{۴}\right)\Big|_{\circ}^{۱/۳}$$

$$= -\log\frac{۱}{۳} + \frac{۱}{۲}\log e = ۲{,}۳۰ \text{ بیت}.$$

(پ) مقداری که در این جا به دست آوردیم کمتر از مقدار به دســـت آمـده در تمریــن (۱.۵- الف) است. این به علّت آن است که چگالی احتمال در این تمرین نقطهٔ اوج دارد که موجب می‌شود عدم‌حتمیت و همراه با آن مقدار اطلاع کاهش یابد.

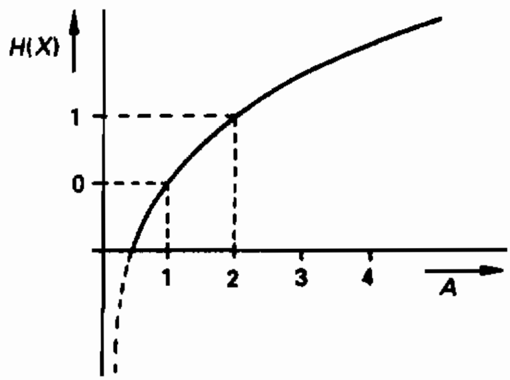
۵.۵ (الف) در این جا از تمرین (۱.۵) نتیجه می‌شود که $H(X) = \log ۲A$. نمودار آن در شکل (۱۸.۵) داده شده است.

(ب) var(x) را می‌توان به طریق زیر محاسبه کرد

$$E(\mathbf{x}) = \int_{-A}^{A} x\,p(x)dx = \int_{-A}^{A}\frac{۱}{۲A}x\,dx = \frac{۱}{۴A}x^{۲}\Big|_{-A}^{A} = \circ.$$

در این صورت با استفاده از این var(x) برابر است با

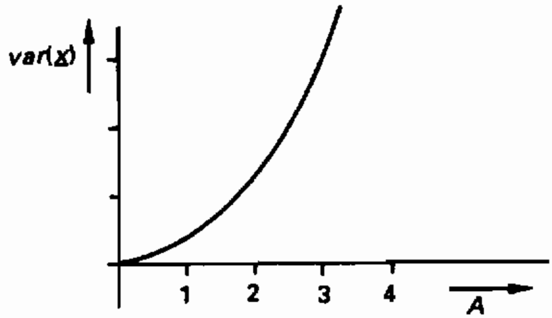
$$\text{var}(\mathbf{x}) = E[(\mathbf{x} - E(\mathbf{x}))^{۲}] = E[\mathbf{x}^{۲}] = \int_{-A}^{A} x^{۲}p(x)\,dx = \frac{۱}{۶A}x^{۳}\Big|_{-A}^{A} = \frac{۱}{۳}A^{۲}.$$



شکل ۱۸.۵- $H(X)$ به صورت تابعی از A (تمرین (۵.۵))

شکل (۱۹.۵) را ببینید.

(پ) A بزرگتر موجب واریانس بزرگتر و عدم‌حتمیت بزرگتر مربوط به $\mathbf{x}$ می‌شود. مطلب اخیر متناظر با افزایش $H(X)$ است همان‌طور که در نمودار $H(X)$ به عنوان تابعی از A نشان داده شد.



شکل ۱۹.۵- واریانس به عنوان تابعی از A (تمرین (۵.۵))

۶.۵ (الف)

$$H(X) = -\int_{0}^{\infty} p(x)\log p(x)\,dx = -\log e\int_{0}^{\infty} p(x)\ln p(x)\,dx$$

$$= -\log e\int_{0}^{\infty}\frac{1}{\lambda}\exp(-\frac{x}{\lambda})\ln[\frac{1}{\lambda}\exp(-\frac{x}{\lambda})]dx$$

$$= \log e\frac{\ln\lambda}{\lambda}\int_{0}^{\infty}\exp(-\frac{x}{\lambda})\,dx + \log e\frac{1}{\lambda^{2}}\int_{0}^{\infty} x\exp(-\frac{x}{\lambda})\,dx$$

$$= \log e\ln\lambda\int_{0}^{\infty} e^{-y}\,dy + \log e\int_{0}^{\infty} ye^{-y}\,dy$$

$$= -\log\lambda e^{-y}\Big|_{0}^{\infty} + \log e\,.\,e^{-y}(-y-1)\Big|_{0}^{\infty} = \log\lambda + \log e.$$

(ب) تابع چگالی احتمال $p(x)$ را باید طوری تعیین کنیم که

$$H(X) = -\int_{0}^{\infty} p(x)\log p(x)\,dx$$

ماکسیمم شود، که در آن، شرایط زیر باید منظور شود:

$$\int_{0}^{\infty} p(x)\,dx = 1 \qquad \text{و} \qquad \int_{0}^{\infty} x\,p(x)\,dx = \lambda$$

اکنون تابع زیر را تشکیل می‌دهیم

$$G(x) = -p(x)\log p(x) + \alpha_1 p(x) + \alpha_2 x\, p(x)$$

اگر مشتق $G(x)$ نسبت به $p(x)$ را برابر صفر قرار دهیم، داریم

$$-\log p(x) - \log e + \alpha_1 + \alpha_2 x = 0.$$

پس از تقسیم بر $\log e$- این عبارت به صورت زیر در می‌آید

$$\ln p(x) + 1 - \lambda_1 - \lambda_2 x = 0,$$

که در آن $\lambda_1 = \dfrac{\alpha_1}{\log e}$ و $\lambda_2 = \dfrac{\alpha_2}{\log e}$. از این، جوابی به صورت زیر به دست می‌آید

$$p(x) = e^{\lambda_1 - 1} e^{\lambda_2 x}.$$

با جایگزین کردن $p(x)$ در هر دو شرط حدّی نتیجه می‌شود

$$\int_0^\infty p(x)\,dx = \int_0^\infty e^{\lambda_1 - 1} e^{\lambda_2 x}\,dx = \frac{e^{\lambda_1 - 1} e^{\lambda_2 x}}{\lambda_2}\Big|_0^\infty = -\frac{e^{\lambda_1 - 1}}{\lambda_2} = 1$$

که در آن فرض شده است که $\lambda_2 < 0$. بنابراین $e^{\lambda_1 - 1} = -\lambda_2$.

همچنین باید داشته باشیم

$$\int_0^\infty x\, p(x)\,dx = \int_0^\infty x e^{\lambda_1 - 1} e^{\lambda_2 x}\,dx = -\int_0^\infty x \lambda_2 e^{\lambda_2 x}\,dx = \lambda,$$

که از آن نتیجه می‌شود که

$$\lambda = -\int_0^\infty x \lambda_2 e^{\lambda_2 x}\,dx = -\lambda_2 \frac{e^{\lambda_2 x}}{(\lambda_2)^2}(\lambda_2 x - 1)\Big|_0^\infty = -\frac{1}{\lambda_2}.$$

بنابراین نتیجه می‌شود که $\lambda_2 = -\dfrac{1}{\lambda}$ و از این $p(x)$ به دست می‌آید

$$p(x) = \frac{1}{\lambda}\exp(-\frac{x}{\lambda})$$

که توزیع نمایی منفی است.

۷.۵ (الف)

$$H(X) = -\int_{-\infty}^{\infty} \frac{1}{2\lambda}\exp(-\frac{|x|}{\lambda})\log[\frac{1}{2\lambda}\exp(-\frac{|x|}{\lambda})]\,dx$$

$$= -\int_{-\infty}^{\infty} \frac{1}{2\lambda}\exp(-\frac{|x|}{\lambda})\log\frac{1}{2}\,dx - \int_{-\infty}^{\infty} \frac{1}{2\lambda}\exp(-\frac{|x|}{\lambda})\log[\frac{1}{\lambda}\exp(-\frac{|x|}{\lambda})]\,dx$$

$$= 1 - \int_{-\infty}^{\infty} \frac{1}{\lambda}\exp(-\frac{x}{\lambda})\log[\frac{1}{\lambda}\exp(-\frac{x}{\lambda})]\,dx.$$

اگر تمرین (۶.۵- الف) را در نظر بگیریم می‌توان نتیجه گرفت که جملهٔ انتگرال‌دار دقیقاً مقدار اطلاع توزیع نمایی منفی است. برای حالتی که اکنون در نظر گرفته‌ایم برای $H(X)$ داریم

$$H(X)=۱+\log\lambda+\log e.$$

(ب) چون $\mathbf{y}=|\mathbf{x}|$ احتمال y برابر احتمال $x-$ و $x+$ خواهد بود. چون $p(x)=p(-x)$ احتمال y دو برابر $p(x)$ خواهد بود. به عبارت دیگر،

$$q(y)=\frac{۱}{\lambda}\exp(-\frac{y}{\lambda}) \qquad , \qquad y\geq ۰$$

اگر این نتیجه را با تمرین (۶.۵- الف) مقایسه کنیم در این صورت مستقیماً نتیجه می‌شود که

$$H(Y)=\log\lambda+\log e.$$

(پ) اگر $\mathbf{x}$ معلوم باشد، $\mathbf{y}$ نیز معلوم است. یعنی، اگر $\mathbf{x}$ معلوم باشد، هیچ عدم‌حتمیتی دربارهٔ $\mathbf{y}$ وجود ندارد و بنابراین داریم $H(Y|X)=۰$.

(ت) به طور کلّی، داریم

$$H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y).$$

اگر مقادیر به دست آمدهٔ قبل برای $H(X)$، $H(Y)$ و $H(Y|X)$ را جایگزین کنیم در این صورت نتیجه می‌شود که

$$H(X|Y)=۱.$$

یعنی، اگر $\mathbf{y}$ معلوم باشد، هنوز هم مقداری عدم‌حتمیت روی $\mathbf{x}$ وجود دارد. واقعیت امر این است چون علامت مجهول است. مقدار $\mathbf{x}$ می‌تواند مثبت یا منفی باشد.

۸.۵ (الف) اگر $x<\frac{۱}{۲}\sqrt{۲}$ آن‌گاه $\mathbf{z}=\mathbf{y}$. بنابراین

$$H(Z|x<\frac{۱}{۲}\sqrt{۲})=H(Y|x<\frac{۱}{۲}\sqrt{۲})=H(Y).$$

برای چگالی احتمال $p(y)$ داریم $p(y)=\sqrt{۲}-y$ برای $۰\leq y\leq\sqrt{۲}$. اکنون مقدار اطلاع را می‌توان به صورت زیر محاسبه کرد

$$H(Y)=-\int_{۰}^{\sqrt{۲}}(\sqrt{۲}-y)\log(\sqrt{۲}-y)dy$$

$$=\int_{\sqrt{۲}}^{۰}t\log t\,dt=\frac{۱}{\ln ۲}t^{۲}(\frac{۱}{۲}\ln t-\frac{۱}{۴})\Big|_{\sqrt{۲}}^{۰}$$

$$= -\log\sqrt{2} + \frac{1}{2\ln 2} = \frac{1}{2}\log\frac{e}{2}$$ بیت.

(ب) چون $z = x$ برای $x \geq \frac{1}{2}\sqrt{2}$، داریم

$$H(Z \mid x \geq \tfrac{1}{2}\sqrt{2}) = H(X \mid x \geq \tfrac{1}{2}\sqrt{2}).$$

از فرمول بیز نتیجه می‌شود که

$$p(x \mid x \geq \tfrac{1}{2}\sqrt{2}) = \frac{p(x)\,.\,p(x \geq \frac{1}{2}\sqrt{2} \mid x)}{p(x \geq \frac{1}{2}\sqrt{2})}$$

$$= \frac{p(x)}{p(x \geq \frac{1}{2}\sqrt{2})} = \frac{4}{3}p(x) \qquad , \qquad \frac{1}{2}\sqrt{2} < x < \sqrt{2}$$

از این عبارت همراه با $p(x) = x$، نتیجه می‌شود که

$$H(X \mid x \geq \tfrac{1}{2}\sqrt{2}) = -\int_{\frac{1}{2}\sqrt{2}}^{\sqrt{2}} \frac{4}{3}p(x)\log\{\tfrac{4}{3}p(x)\}\,dx = -\int_{\frac{1}{2}\sqrt{2}}^{\sqrt{2}} \frac{4}{3}x\log\{\tfrac{4}{3}x\}\,dx$$

$$= -\int_{\frac{2}{3}\sqrt{2}}^{\frac{4}{3}\sqrt{2}} \frac{3}{4}t\log t\,dt = -\frac{3}{4\ln 2}t^2(\tfrac{1}{2}\ln t - \tfrac{1}{4})\Bigg|_{\frac{2}{3}\sqrt{2}}^{\frac{4}{3}\sqrt{2}}$$

$$= -\frac{8}{3}[\tfrac{1}{2}\log(4\sqrt{2}) - \tfrac{1}{2}\log 3 - \tfrac{1}{4}\log e] + \frac{2}{3}[\tfrac{1}{2}\log(2\sqrt{2}) - \tfrac{1}{2}\log 3 - \tfrac{1}{4}\log e]$$

$$= -\frac{16}{7} + \log(3\sqrt{e})$$ بیت.

(پ) $$H(Z) = p(x < \tfrac{1}{2}\sqrt{2})\,.\,H(Z \mid x < \tfrac{1}{2}\sqrt{2}) + p(x \geq \tfrac{1}{2}\sqrt{2})\,.\,H(Z \mid x > \tfrac{1}{2}\sqrt{2}).$$

چون $p(x < \frac{1}{2}\sqrt{2}) = \frac{1}{4}$ و $p(x \geq \frac{1}{2}\sqrt{2}) = \frac{3}{4}$، همراه با مقادیر بـه دسـت آمـدهٔ قبلی در (الف) و (ب) نتیجه می‌شود که

$$H(Z) = \frac{1}{4} \times \frac{1}{2}\log\frac{e}{2} + \frac{3}{4}[-\frac{17}{6} + \log(3\sqrt{e})] = -\frac{1}{2}\log e - \frac{9}{4} + \frac{3}{4}\log 3.$$

۹.۵ (الف) برای $p(x,y)$ داریم

$$p(x,y) = \frac{1}{32} \qquad , \qquad 0 \leq x \leq 8 \quad \text{و} \quad 0 \leq y \leq 8 - x$$

چگالی احتمال $p(x)$ را می‌توان به صورت زیر محاسبه کرد

$$p(x)=\int_{۰}^{۸-x}p(x,y)\,dy=\frac{۸-x}{۳۲}.$$

در این صورت مقدار اطلاع $H(X)$ برابر است با

$$H(X)=-\int_{۰}^{۸}\frac{۸-x}{۳۲}\log(\frac{۸-x}{۳۲})dx=\int_{\frac{۱}{۴}}^{۰}۳۲t\log t\,dt$$

$$=۳۲\frac{t^{۲}}{\ln ۲}(\frac{۱}{۲}\ln t-\frac{۱}{۴})\Big|_{۱/۴}^{۰}=۲+\frac{۱}{۲}\log e.$$

(ب) داریم $p(y|x)=\frac{p(x,y)}{p(x)}=\frac{۱}{۸-x}$، که در آن $۰<y<۸-x$. اکنون برای مقدار اطلاع شرطی نتیجه می‌شود که

$$H(Y|X)=-\int_{۰}^{۸}\int_{۰}^{۸-x}p(x,y)\log p(y|x)\,dy\,dx$$

$$=\int_{۰}^{۸}\int_{۰}^{۸-x}\frac{۱}{۳۲}\log(۸-x)\,dy\,dx=\frac{۱}{۳۲}\int_{۰}^{۸}(۸-x)\log(۸-x)\,dx$$

$$=-\frac{۱}{۳۲}\int_{۸}^{۰}t\log t\,dt=-\frac{۱}{۳۲}\frac{t^{۲}}{\ln ۲}(\frac{۱}{۲}\ln t-\frac{۱}{۴})\Big|_{۸}^{۰}=۳-\frac{۱}{۲}\log e.$$

(پ) با کمک جوابهای (الف) و (ب) نتیجه می‌شود که

$$H(X,Y)=H(X)+H(Y|X)=۲+\frac{۱}{۲}\log e+۳-\frac{۱}{۲}\log e=۵\ \text{بیت}.$$

با محاسبهٔ مستقیم به دست می‌آوریم

$$H(X,Y)=-\int_{۰}^{۸}\int_{۰}^{۸-x}p(x,y).\log p(x,y)\,dx\,dy=\int_{۰}^{۸}\int_{۰}^{۸-x}\frac{۱}{۳۲}\log\frac{۱}{۳۲}\,dx\,dy=۵\ \text{بیت}.$$

۵. ۱۰ (الف) چون $p(x,y)$ چگالی احتمال گاوسی دو-بعدی است که در آن $E(\mathbf{y})=۰$ و $\mathrm{var}(\mathbf{y})=\sigma_{۲}^{۲}$، برای $q(y)$ نتیجه می‌شود که

$$q(y)=\frac{۱}{\sigma_{۲}\sqrt{۲\pi}}\exp(-\frac{y^{۲}}{۲\sigma_{۲}^{۲}})$$

برمبنای قضیهٔ (۳.۵) داریم

$$H(Y)=\log(\sigma_{۲}\sqrt{۲\pi e}).$$

(ب) چگالی احتمال $p(x|y)$ را می‌توان به صورت زیر به دست آورد

$$p(x|y)=\frac{p(x,y)}{q(y)}$$

$$=\frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{x^2}{\sigma_1^2}-\frac{2\rho xy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right\}+\frac{y^2}{2\sigma_2^2}\right]$$

$$=\frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left[-\frac{(x-\frac{\sigma_1}{\sigma_2}\rho y)^2}{2\sigma_1^2(1-\rho^2)}\right]$$

این نیز یک چگالی احتمال گاوسی است با میانگین شرطی $(\frac{\sigma_1}{\sigma_2})\rho y$ و واریانس $\sigma_1^2(1-\rho)^2$.

اکنون برای $H(X|Y)$ نتیجه می‌شود که

$$-\int_{-\infty}^{\infty}q(y)\left\{\int_{-\infty}^{\infty}p(x|y)\log p(x|y)dx\right\}dy,$$

که در آن $p(x|y)$ برابر چگالی احتمال گاوسی داده شده در بالاست. چون مقدار اطلاع در حالت چگالی احتمال گاوسی تنها به واریانس بستگی دارد و به میانگین بستگی ندارد، که به سادگی می‌توان نشان داد، با کمک قضیهٔ (۳.۵) نتیجه می‌شود که

$$-\int_{-\infty}^{\infty}p(x|y)\log p(x|y)dy=\log[\sigma_1\sqrt{1-\rho^2}\sqrt{2\pi e}],$$

و همچنین داریم

$$H(X|Y)=\log[\sigma_1\sqrt{1-\rho^2}\sqrt{2\pi e}].$$

(پ) چون $H(X,Y)=H(Y)+H(X|Y)$ از نتایج (الف) و (ب) نتیجه می‌شود که

$$H(X,Y)=\log[\sigma_2\sqrt{2\pi e}]+\log[\sigma_1\sqrt{1-\rho^2}\sqrt{2\pi e}]$$

$$=\log[\sigma_1\sigma_2\sqrt{1-\rho^2}\,2\pi e].$$

(ت) اگر $\rho=0$ در این صورت این بدین معناست که $(\mathbf{x}-E(\mathbf{x}))$ و $(\mathbf{y}-E(\mathbf{y}))$ هیچ گونه وابستگی نشان نمی‌دهند. چون می‌دانیم که $E(\mathbf{x})=E(\mathbf{y})=0$ بنابراین نتیجه می‌شود که $\mathbf{x}$ و $\mathbf{y}$ هیچ وابستگی ندارند. به عبارت دیگر $\mathbf{x}$ و $\mathbf{y}$ از یکدیگر مستقلند. با

جایگزینی $\rho=0$ در $H(X|Y)$ نتیجه می‌شود

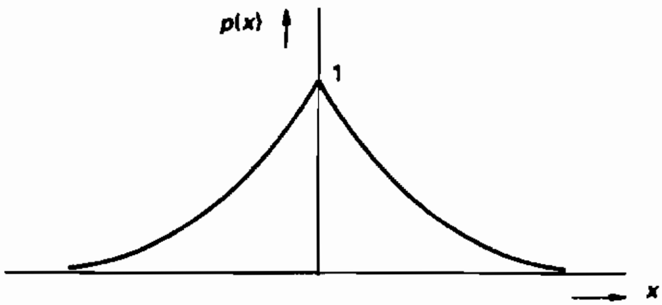
$$H(X|Y)=\log[\sigma_1\sqrt{2\pi e}]$$

که با $H(X)$ یکی است؛ که با این حقیقت که برای $\rho=0$، **x** و **y** مستقلّند سازگار است. با جایگزینی $\rho=0$ در $H(X,Y)$ همچنان که در (پ) محاسبه شد حاصل می‌شود

$$H(X,Y)=\log[\sigma_1\sigma_2 2\pi e]=H(X)+H(Y),$$

که مجدداً با استقلال **x** و **y** که با $\rho=0$ حاصل می‌شود مطابقت دارد.

۱۱.۵ (الف) تابع چگالی احتمال برابر $p(x)=e^{-a|x|}$ است (شکل (۵. ۲۰) را ببینید). ثابت a با انتگرال‌گیری و برابر با یک قرار دادن انتگرال به دست می‌آید؛ که پی‌درپی نتیجه می‌دهد



شکل ۵. ۲۰- چگالی احتمال تمرین (۱۱.۵)

$$\int_{-\infty}^{0}e^{ax}\,dx+\int_{0}^{\infty}e^{-ax}\,dx=1,$$

$$\rightarrow\qquad 2\int_{0}^{\infty}e^{-ax}\,dx=1,$$

$$\rightarrow\qquad -\frac{2}{a}e^{-ax}\Big|_{0}^{\infty}=1,$$

به قسمی که $a=2$ و $p(x)=e^{-2|x|}$.

برای $H(X)$ بنابر تعریف داریم

$$H(X)=-\int_{-\infty}^{\infty}p(x)\log p(x)\,dx.$$

پس از جایگزینی $p(x)$ نتیجه می‌شود که

$$H(X) = -\int_{-\infty}^{۰} e^{۲x} \log e^{۲x}\, dx - \int_{۰}^{\infty} e^{-۲x} \log e^{-۲x}\, dx.$$

با بهره‌وری از تقارن $p(x)$ نتیجه می‌شود که

$$H(X) = ۴ \log e \int_{۰}^{\infty} x e^{-۲x}\, dx$$

$$= -۲\log e \int_{۰}^{\infty} x\, de^{-۲x} = -۲x\, e^{-۲x} \log e \Big|_{۰}^{\infty} + ۲\log e \int_{۰}^{\infty} e^{-۲x}\, dx$$

$$= -e^{-۲x} \log e \Big|_{۰}^{\infty} = \log e$$ بیت / نماد.

(ب) از توان اطلاع P_H یک سیگنال تصادفی $\mathbf{x}(t)$ تولید شده توسط یک منبع اطلاع می‌فهمیم که توان یک سیگنال گاوسی مقدار اطلاعی به بزرگی سیگنال تصادفی $\mathbf{x}(t)$ دارد.

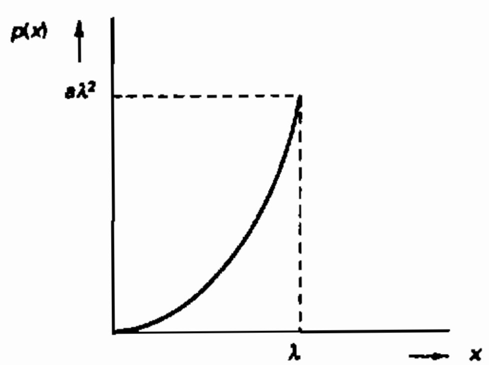
(پ) توان اطلاع با عبارت زیر داده شده است:

$$P_H = \frac{۱}{۲\pi e} ۲^{۲H(X)}.$$

با جایگزینی $H(X)$ نتیجه می‌شود

$$P_H = \frac{۱}{۲\pi e} ۲^{۲\log e} = \frac{e^۲}{۲\pi e} = \frac{e}{۲\pi}.$$

۱۲.۵ (الف) شکل (۲۱.۵) را ببینید.



شکل ۲۱.۵- چگالی احتمال تمرین (۱۲.۵)

نخست پارامتر a را با محاسبهٔ انتگرال و مساوی یک قرار دادن آن، به دست می‌آوریم

$$\int_0^\lambda ax^2\,dx = \frac{1}{3}ax^3\Big|_0^\lambda = \frac{1}{3}a\lambda^3 = 1,$$

بنابراین

$$a = \frac{3}{\lambda^3} \qquad \text{و} \qquad p(x) = \frac{3}{\lambda^3}x^2.$$

اکنون مقدار اطلاع در یک نمونه برابر است با

$$H(X) = -\int_0^\lambda \frac{3}{\lambda^3}x^2\log\left(\frac{3}{\lambda^3}x^2\right)dx = -\frac{6}{\lambda^3}\int_0^\lambda x^2 . \log x\,dx - \frac{3}{\lambda^3}\int_0^\lambda x^2\log\frac{3}{\lambda^3}\,dx$$

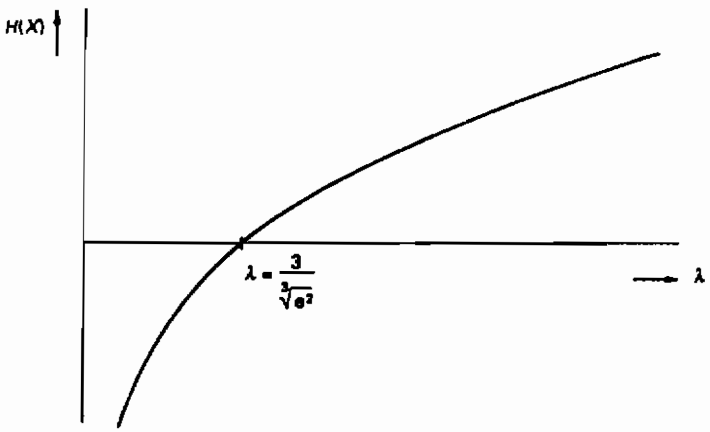
$$= -\frac{2}{\lambda^3}\int_0^\lambda \log x\,dx^3 - \frac{3}{\lambda^3}\log\frac{3}{\lambda^3}\times\frac{1}{3}\lambda^3$$

$$= -\frac{2}{\lambda^3}x^3\log x\Big|_0^\lambda + \frac{2}{\lambda^3}\int_0^\lambda x^3\,d\log x - \log\frac{3}{\lambda^3}$$

$$= -2\log\lambda + \frac{2}{\lambda^3}\log e\times\frac{1}{3}x^3\Big|_0^\lambda - \log\frac{3}{\lambda^3} = -2\log\lambda + \frac{2}{3}\log e - \log\frac{3}{\lambda^3}$$

$$= \log\lambda - \log 3 + \frac{2}{3}\log e = \log\frac{\lambda e^{\frac{2}{3}}}{3}\ \text{بیت / نماد}.$$

شکل (۲۲.۵) را ببینید.



شکل ۲۲.۵- $H(X)$ به عنوان تابعی از λ (تمرین (۱۲.۵))

(ب) از عبارت به دست آمده در بند (الف) دیده می‌شود که $H(X)=0$، برای λ داریم

$$\lambda = 3e^{-1/3}.$$

$H(X)$ برای مقادیر کمتر λ منفی می‌شود.

از این رو $\lambda = 3e^{-2/3}$ را می‌توان به عنوان کران پایین در نظر گرفت که هنوز هم برای کار کردن با دقّت $\Delta x = 1$ معنی‌دار است.

(پ) توان اطلاع با عبارت زیر داده شده است:

$$P_H = \frac{1}{2\pi e} 2^{2H(X)}.$$

با جایگزینی $H(X)$ نتیجه می‌شود

$$P_H = \frac{1}{2\pi e} 2^{2\log(\lambda e^{1/2}/3)} = \frac{1}{2\pi e}\left(\frac{\lambda e^{1/2}}{3}\right)^2.$$

۱۳.۵ (الف) شکل (۲۳.۵) را ببینید.

برای $H(X)$ بنابر تعریف داریم

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx$$

پس از جایگزینی $p(x)$ نتیجه می‌شود که

$$H(X) = -\frac{1}{a}\int_{-a}^{0}(1+\frac{x}{a})\log\{\frac{1}{a}(1+\frac{x}{a})\}\,dx - \frac{1}{a}\int_{0}^{a}(1-\frac{x}{a})\log\{\frac{1}{a}(1-\frac{x}{a})\}\,dx.$$

با استفاده از تقارن $p(x)$ نتیجه می‌شود که

$$H(X) = -\frac{2}{a}\int_{0}^{a}(1-\frac{x}{a})\log\{\frac{1}{a}(1-\frac{x}{a})\}\,dx$$

$$= -\frac{2}{a}\int_{0}^{a}(1-\frac{x}{a})\log\frac{1}{a}\,dx - \frac{2}{a}\int_{0}^{a}(1-\frac{x}{a})\log(1-\frac{x}{a})\,dx.$$

اکنون برای اوّلین انتگرال پس از علامت تساوی نتیجه می‌شود

$$= -2\log a\int_{0}^{a}(1-\frac{x}{a})\,d(1-\frac{x}{a}) = -2\log a \times \frac{1}{2}(1-\frac{x}{a})^2\Big|_{0}^{a}$$

$$= -2\log a - \frac{1}{2} = \log a.$$

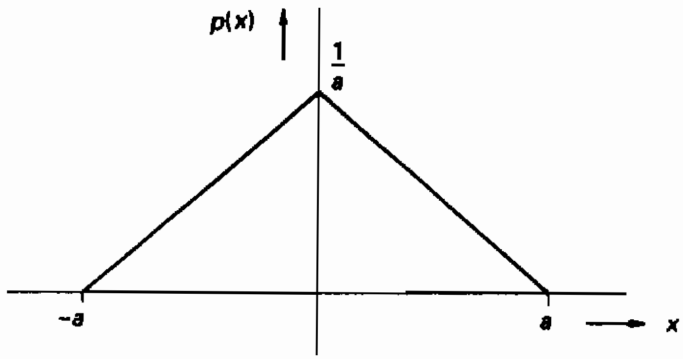
برای دومین انتگرال پس از علامت تساوی پس از جایگزینی $(1-\frac{x}{a})=e^{-z}$ نتیجه می‌شود که

$$=2\int_0^a(1-\frac{x}{a})\log(1-\frac{x}{a})d(1-\frac{x}{a})=2\int_0^a e^{-z}\log e^{-z}\,de^{-z}$$

$$=2\log e\int_0^\infty z\,.\,e^{-2z}\,dz=\frac{1}{2}\log e,$$

بنابراین $H(X)$ برابر است با

$$H(X)=\log a+\frac{1}{2}\log e=\log a\sqrt{e}\ \text{بیت / نماد}.$$



شکل ۵.۲۳- چگالی احتمال تمرین (۵.۱۳)

(ب) توان اطلاع منبع به صورت زیر تعریف می‌شود

$$P_H=\frac{1}{2\pi e}2^{2H(X)}.$$

با جایگزینی $H(X)$ نتیجه می‌شود که

$$P_H=\frac{1}{2\pi e\times 2^{2\log(a\sqrt{e})}}=\frac{1}{2\pi e}2^{\log(a^2 e)}=\frac{a^2 e}{2\pi e}=\frac{a^2}{2\pi}.$$

(پ) توان P_f به صورت زیر تعریف می‌شود

$$P_f=\mathrm{var}(\mathbf{x})=\int_{-\infty}^{\infty}\{x-E(\mathbf{x})\}^2\,p(x)dx.$$

چون $E(\mathbf{x})=0$، این عبارت را می‌توان به صورت زیر نوشت

$$P_f=\int_{-\infty}^{\infty}x^2\,p(x)dx=\frac{2}{a}\int_0^a x^2(1-\frac{x}{a})dx=\frac{a^2}{6}.$$

(ت) برای مقدار اطلاع $H(Y)$ یک نمونه از سیگنال گاوسی داریم

$$H(Y) = \log \sigma \sqrt{۲\pi e}.$$

زیرا منبع گاوسی باید توانی برابر با منبعی که $\sigma = \sqrt{\frac{a^۲}{۶}}$ را در نظر گرفته است باشد؛ بنابراین داریم

$$H(Y) = \log \sqrt{\frac{۲\pi e a^۲}{۶}}.$$

چون $H(X) = \log a\sqrt{e}$ با مقایسهٔ $H(X)$ و $H(Y)$ نتیجه می‌شود

$$H(Y) - H(X) = \log \sqrt{\frac{۲\pi e a^۲}{۶ e a^۲}} = \frac{۱}{۲} \log \frac{\pi}{۳} = ۰٫۰۳۴.$$

به عبارت دیگر، محتوای اطلاع $H(X)$ منبع داده شده از محتوای اطلاع $H(Y)$ منبع گاوسی با توان مساوی برای هر دو منبع کوچکتر است.

۶

کانال ارتباطی پیوسته

۱.۶ ظرفیت کانالهای ارتباطی پیوسته

رمزنگاری و رمزگشایی در کانال ارتباطی پیوسته با حالت کانال گسسته تفسیر متفاوتی دارد. حال برای رمزنگاری می‌توان به عنوان مثال با روشهای مدولاسیون مثل مدولاسیون دامنهٔ نوسان، مدولاسیون فرکانس یا کراندار کردن پهنای باند مثلاً از راه به‌کارگیری صافیهای باند با عبور کم توجّه کرد. به این طریق سیگنال تصادفی تولید شده توسط منبع و دارای اطلاع را به شکل مناسبی برای کانال پیوسته تبدیل می‌کنیم.

به طور اجتناب ناپذیری نوفه با سیگنال در کانال ارسال اضافه می‌شود. در طرف دریافت‌کننده باید سیگنال ارسال شده از سیگنال دریافت شده بازسازی شود. ظرفیت کانال نیز برای الگوی ارتباطی پیوسته نقش مهمی را ایفا می‌نماید، اگر بخواهیم تعیین کنیم که تحت چه شرایطی ماکسیمم مقدار اطلاع را می‌توان ارسال کرد.

کانال پیوسته‌ای در نظر می‌گیریم که در آن یک سیگنال پیوسته $\mathbf{x}(t)$ در ورودی پیشنهاد شده است و در خروجی سیگنال پیوسته $\mathbf{y}(t)$ به دست می‌آید. مجدداً N نمونه از این سیگنالها را در نظر می‌گیریم. اکنون تابع چگالی احتمال سیگنال دریافت شده $y(t)$ برای سیگنال ارسال شدهٔ معلوم $x(t)$ را می‌توان به صورت زیر نوشت

$$q(\tilde{y}\,|\,\tilde{x}) = q(y_1,\dots,y_N\,|\,x_1,\dots,x_N).$$

اگر فرض کنیم که یک نمونه $y_i = y(t_i)$ تنها به یک نمونه $x_i = x(t_i)$ وابسته است، می‌توان از یک کانال پیوستهٔ بی‌حافظه صحبت کرد. در این حالت داریم

$$q(\tilde{y}|\tilde{x})=\prod_{i=1}^{N}q(y_i|x_i). \qquad (۱.۶)$$

بنابراین اکنون می‌توان ویژگیهای کانال را برای یک زوج تکی در نمونه‌های x و y، که تجزیه و تحلیل را به طور قابل ملاحظه‌ای ساده‌تر می‌کند تعیین کرد. به علاوه مقدار اطلاع $H(X)$ و $H(Y)$ به ترتیب متعلّق به طرف ارسال شده و طرف دریافت شده‌اند، تعدادی دیگر کمیتهای مهم مانند $H(X,Y)$، $H(X|Y)$ و $H(Y|X)$ وجود دارند. کمیت مهم دیگر اطلاع متقابل $I(X;Y)=I(Y;X)$ است، که نشان می‌دهد x چقدر اطلاع دربارهٔ y می‌دهد یا y چقدر اطلاع دربارهٔ x می‌دهد.

برای دو متغیر تصادفی پیوسته **x** و **y**، که نمونه‌های سیگنالهای $x(t)$ و $y(t)$ را به ترتیب در طرف ارسال‌کننده و دریافت‌کننده با چگالی احتمال توأم $p(x,y)$ نشان می‌دهند اطلاع متقابل $I(X;Y)$ به صورت زیر تعریف می‌شود

$$I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\log\frac{p(x,y)}{p(x)\cdot q(y)}dx\,dy, \qquad (۲.۶)$$

که متناظر با معادلهٔ (۵۵.۵) است. با وجود این، در این جا با تفسیر خاصّی از معادلهٔ (۵۵.۵) که **x** و **y** را به ترتیب به کانال ورودی و خروجی نسبت می‌دهد سروکار داریم.

اکنون رابطه‌های زیر را برای مقادیر گوناگون اطلاع به کار می‌بریم

$$\begin{aligned}I(X;Y)&=H(X)+H(Y)-H(X,Y)\\&=H(X)-H(X|Y)\\&=H(Y)-H(Y|X).\end{aligned} \qquad (۳.۶)$$

چون $H(Y|X)\le H(Y)$، بلافاصله می‌توان دید که

$$I(X;Y)\ge 0, \qquad (۴.۶)$$

اگر **x** و **y** به طور آماری مستقل باشند تساوی برقرار می‌شود. در حالتی که **x** و **y** یکسانند همچنین می‌توان نتیجه گرفت که $I(X;X)=H(X)$. اغلب اطلاع متقابل $I(X;Y)$ با نرخ ارسال R نیز نشان داده می‌شود چون مقدار اطلاع ارسال شده روی کانال را نشان می‌دهد.

درست مانند موردی که در بخش (۳.۵) داشتیم اطلاع پیوسته را فقط می‌توان تحت قیدهای فرض شده تعیین کرد، قیدها باید روی توزیع احتمال ورودی یک کانال پیوسته فرض شوند. در غیراین صورت، ورودی کانال می‌تواند هر عددی روی خطّ حقیقی

نامحدود باشد، که از نظر فیزیکی موقعیت غیرممکنی است.

ظرفیت C یک کانال پیوسته ماکسیمم نرخ ارسال R یا $I(X;Y)$ است که می‌توان آن را با اتصال همهٔ منابع اطلاع ممکن، سازگار با قیدها، به کانال به دست آورد.

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} \{H(Y) - H(Y|X)\}. \qquad (۵.۶)$$

به طور کلّی، محاسبهٔ ظرفیت یک کانال پیوسته کار مشکلی است. تنها برای برخی از کانالهای خاص، مثلاً کانالهای با نوفه گاوسی سفید جمعی با محدودیّت توان متوسط، ممکن است یک عبارت تحلیلی برای ظرفیت به دست آورد. در حالتهای دیگر روشهای عددی را باید به کار برد. نخست، نوع عمومی موسوم به کانالهای جمعی را که در آن نوفه به سیگنال ارسال شدهٔ **x** اضافه شدهٔ و به طور آماری مستقل از آن است، بررسی می‌کنیم. برای یک مقدار داده شدهٔ x اگر $n = y - x$ برای نوفه برقرار باشد، می‌توان مقدار y را به دست آورد، بنابراین باید داشته باشیم

$$q(y|x) = q(x+n|x) = p(n|x). \qquad (۶.۶)$$

چون نوفه از سیگنال ورودی مستقل است، نتیجه می‌شود که:

$$q(y|x) = p(n|x) = p(n) = p(y-x). \qquad (۷.۶)$$

برای کانال جمعی برای اثر نوفه داریم

$$\begin{aligned} H(Y|X) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log q(y|x)\,dx\,dy \\ &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)q(y|x)\log q(y|x)\,dx\,dy \\ &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)p(n)\log p(n)\,dx\,dn \\ &= -\int_{-\infty}^{\infty} p(n)\log p(n)\,dn = H(\mathrm{N}). \end{aligned} \qquad (۸.۶)$$

در این حالت، برای ظرفیت به دست می‌آوریم

$$C = \max_{p(x)} \{H(Y) - H(\mathrm{N})\} = \max_{p(x)} \{H(Y)\} - H(\mathrm{N}). \qquad (۹.۶)$$

از این رو، برای ماکسیمم کردن R لازم است $H(Y)$ را ماکسیمم کنیم. امّا این ماکسیمم مقیّد به سیگنالهای ارسال شده‌ای است که باید مورد توجّه قرار گیرند.

مثال ۱.۶

توابع توزیع احتمال نسبت به **x** و **n** عبارتند از:

$$p(x)=\frac{۱}{۸} \quad , \quad -۴\le x\le ۴,$$
$$=۰ \quad , \quad \text{برای سایر نقاط.}$$

$$p(n)=\frac{۱}{۲} \quad , \quad -۱\le n\le ۱,$$
$$=۰ \quad , \quad \text{برای سایر نقاط.}$$

توزیع احتمال توأم $p(x,y)$ را می‌توان محاسبه کرد، که در آن $\mathbf{y}=\mathbf{x}+\mathbf{n}$.

$$p(n)=q(y|\,x)=\frac{p(x,y)}{p(x)}.$$

بنابراین، شکل (۱.۶) را ملاحظه کنید؛

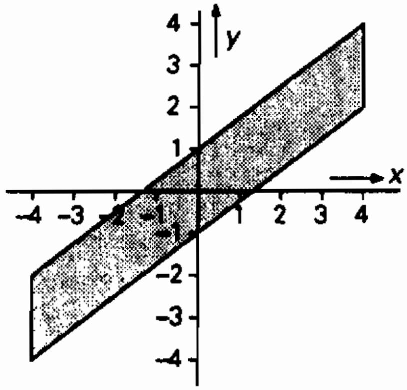
$$p(x,y)=\begin{cases}\frac{۱}{۱۶} & , \quad -۴\le x\le ۴ \quad \text{و} \quad -۱\le y-x\le ۱,\\ ۰ & , \quad \text{برای سایر نقاط.}\end{cases}$$

توزیعهای احتمال $q(y)$ و $p(x|\,y)$ را می‌توان به صورت زیر تعیین کرد (شکل (۲.۶) را ببینید):

$$q(y)=\begin{cases}\int_{-۴}^{y+۱}\frac{۱}{۱۶}dx=\frac{۱}{۱۶}(y+۵) & , \quad -۵\le y\le -۳,\\ \int_{y-۱}^{y+۱}\frac{۱}{۱۶}dx=\frac{۱}{۸} & , \quad -۳\le y\le ۳,\\ \int_{y-۱}^{۴}\frac{۱}{۱۶}dx=\frac{۱}{۱۶}(۵-y) & , \quad ۳\le y\le ۵,\end{cases}$$

$$p(x|\,y)=\frac{p(x,y)}{q(y)}=\begin{cases}\frac{۱}{y+۵} & , \quad -۵\le y\le -۳ \quad , \quad -۴\le x\le y+۱,\\ \frac{۱}{۲} & , \quad -۳\le y\le ۳ \quad , \quad y-۱\le x\le y+۱,\\ \frac{۱}{۵-y} & , \quad ۳\le y\le ۵ \quad , \quad y-۱\le x\le ۴.\end{cases}$$

با کمک توزیع احتمال اندازه‌های گوناگونی را می‌توان محاسبه کرد.



شکل ۶.۱- دامنهٔ تعریف $p(x,y)$ مربوط به مثال (۶.۱)

$$H(X)=۳,$$

$$H(Y|X)=H(\mathrm{N})=۱,$$

$$H(X,Y)=۴,$$

$$H(Y)=۳+\frac{۱}{۸}\log e,$$

$$H(X|Y)=۱-\frac{۱}{۸}\log e.$$

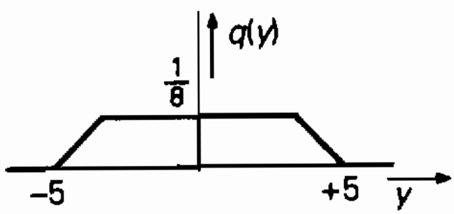
از این رو نرخ اطلاع برابر است با

$$I(X;Y)=H(Y)-H(\mathrm{N})=۲+\frac{۱}{۸}\log e.$$

ظرفیت را می‌توان با تغییر دادن ورودی به قسمی که $H(Y)$ را ماکسیمم کند محاسبه کرد.

$$C=\lim_{p(x)} H(Y)-۱.$$

این کار تنها اگر قیدی روی ماکسیمم قبلاً تعیین شده باشد امکان‌پذیر است. ▲



شکل ۶.۲- چگالی احتمال $q(y)$ مربوط به مثال (۶.۱)

در حالت پیوسته عموماً مناسب نیست فرض کنیم که کانال بی‌حافظه است. در این حالت نرخ ارسال را به صورت زیر تعریف می‌کنیم

$$I(\tilde{X};\tilde{Y}) = \lim_{N\to\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\tilde{x},\tilde{y}) \log \frac{p(\tilde{x},\tilde{y})}{p(\tilde{x})\cdot q(\tilde{y})} d\tilde{x}\, d\tilde{y} \qquad (۱۰.۶)$$

که در آن $\tilde{X}$ و $\tilde{Y}$ بردارهای تصادفی N- بعدی هستند. در این صورت ظرفیت یک کانال باحافظه برابر است با

$$C = \lim_{N\to\infty} \max_{p(\tilde{x})} I(\tilde{X};\tilde{Y}). \qquad (۱۱.۶)$$

ظرفیت کانالهای باحافظه در بخش (۵.۶) بررسی خواهد شد.

۲.۶ ظرفیت در حالت نوفه سفید گاوسی جمعی

سیگنالی که باید منتقل شود و نوفه اضافه شده در کانال اغلب با توانشان مشخّص می‌شوند. فرض می‌کنیم که نوفه جمعی است و مستقل از سیگنال ارسال شده $x(t)$ می‌باشد. در بسیاری از حالات این یک فرض واقعی است. در این صورت برای توان (متوسط) P_y در طرف دریافت‌کننده نتیجه می‌شود که

$$P_y = P_x + P_n, \qquad (۱۲.۶)$$

که در آن P_n متوسط توان نوفه است. به علاوه فرض می‌کنیم که نوفه ***سفید*** است، یعنی، دارای چگالی طیفی توان مسطح بر پهنای باند W $(W\to\infty)$ است. علاوه بر این فرض می‌کنیم که مقدارش (دامنهٔ نوسان) در هر زمان با یک تابع چگالی احتمال گاوسی $N(\circ,\sigma^{۲})$ مشخّص می‌گردد. در این حالت از ***نوفه سفید گاوسی*** صحبت می‌کنیم. برای محاسبهٔ تابع همبستگی برای نوفه سفید، با به کارگیری معادلهٔ (۳۹.۵)، طیف مسطح یک تابع دلتا یا قلّه‌دار $R_{xx}(\tau)=\circ$ برای $\tau\neq\circ$ حاصل می‌گردد. در نتیجه نوفه سفید گاوسی دارای هیچ‌گونه وابستگی درونی نبوده و نمونه‌ها به طور تصادفی مستقل هستند. بنابراین می‌توان نمونه‌ها را به طور جداگانه بررسی کرد.

می‌خواهیم ظرفیت را برحسب بیت بر ثانیه بیان کنیم و همچنین به عنوان یک نتیجه $H(Y)$ و $H(N)$ را برحسب بیت بر ثانیه خواهیم داد. با به‌کارگیری نتایج بخش (۳.۵) نتیجه می‌شود که مقدار اطلاع سیگنال نوفه برابر است با

$$H(N) = \log \sigma_n \sqrt{۲\pi e} \quad \text{نمونه / بیت}$$

$$= W \log\{2\pi e \sigma_n^2\} \text{ ثانیه / بیت.} \qquad (۱۳.۶)$$

برای سیگنالی که به توان متوسط معیّنی محدود است این ماکسیمم مقدار اطلاع نیز می‌باشد. این بدین معناست که چون ظرفیت کانال پیوسته با افزایش $H(\mathrm{N})$ کاهش می‌یابد، این نوفه گاوسی متناظر با موقعیت بدترین حالت می‌باشد. به این علت، محاسبهٔ ظرفیت در حالت نوفه گاوسی به طور کلّی یک کران پایین برای ظرفیت واقعی خواهد داد، که از نقطه نظر طراح کاملاً پذیرفتنی است.

برای ظرفیت یک کانال ارتباطی پیوسته با نوفهٔ جمعی داریم (معادلهٔ (۹.۶))

$$C = \max_{p(x)} \{H(Y)\} - H(\mathrm{N}) \text{ ثانیه / بیت.}$$

ماکسیمم $H(Y)$ با $\mathbf{y} = \mathbf{x} + \mathbf{n}$ در حالتی که $\mathbf{y}$ دارای توزیع گاوسی است با توان $P_y = \sigma_y^2$ رخ خواهد داد. چون نوفه گاوسی است نتیجه خواهد شد که سیگنال ورودی نیز باید گاوسی باشد.

اکنون چون $P_y = \sigma_y^2 = \sigma_x^2 + \sigma_n^2$ بنابراین نتیجه می‌شود که:

$$\max_{p(x)} H(Y) = \frac{1}{2} \log\{2\pi e(\sigma_x^2 + \sigma_n^2)\} \text{ نمونه / بیت}$$

$$= W \log\{2\pi e(\sigma_x^2 + \sigma_n^2)\} \text{ ثانیه / بیت.} \qquad (۱۴.۶)$$

بنابراین ظرفیت کانال در حالتی که توان متوسط مقیّد است برابر است با

$$C = W \log\{2\pi e(\sigma_x^2 + \sigma_n^2)\} - W \log\{2\pi e \sigma_n^2\}$$

$$= W \log\left\{\frac{\sigma_x^2 + \sigma_n^2}{\sigma_n^2}\right\} = W \log\left\{\frac{P_x + P_n}{P_n}\right\}$$

$$= W \log\left\{1 + \frac{P_x}{p_n}\right\} \text{ ثانیه / بیت.} \qquad (۱۵.۶)$$

همچنان که دیده می‌شود ارسال با نرخی بالاتر از C بدون معرّفی خطاها امکان‌پذیر نیست.

از عبارت یافته شده برای ظرفیت کانال برمی‌آید که می‌توان تعادلی بین پهنای باند W و نسبت سیگنال به نوفه $\frac{P_x}{P_n}$ برقرار کرد و هنوز هم همان ظرفیت را حفظ نمود.

۳.۶ کرانهای ظرفیت در حالت نوفه سفید غیرگاوسی

برای ظرفیت کانال داده شده بدترین وضعیت ممکن را در نظر می‌گیریم، یعنی یک سیگنال نوفهٔ گاوسی که دارای ماکسیمم مقدار اطلاع است، در صورتی که طیف چگالی توان مسطح فرض می‌شود به قسمی که نمونه‌ها به طور آماری مستقل باشند (کانال بی‌حافظه). اگر این شرایط برقرار نباشد مقدار متفاوتی برای ظرفیت به دست خواهد آمد. اگر نوفه گاوسی نباشد ظرفیت بزرگتر خواهد شد. با وجود این، تعیین تحلیلی ظرفیت کانال معمولاً پیچیده خواهد بود. با روشهای عددی موجود می‌توان ظرفیت را با تقریب خوبی تعیین کرد. همچنین می‌توان کرانهای بالا و پایینی برای ظرفیت به دست آورد که تعیین آنها نسبتاً ساده است. در این صورت مفهوم توان اطلاع P_H را به کار می‌گیریم؛ وضعیتی که در آن نوفه واقعاً نسبت به توان کراندار ولی غیرگاوسی است در نظر می‌گیریم. به علاوه فرض می‌کنیم که نمونه‌ها به طور آماری مستقلّند. مقدار اطلاع نوفه برابر است با (با معادلهٔ (۷۳.۵) مقایسه کنید)

$$H(\mathrm{N}) = \log\sqrt{2\pi e P_{H_n}} \text{ نمونه / بیت}$$

$$= W\log\{2\pi e P_{H_n}\} \text{ ثانیه / بیت.} \qquad (۱۶.۶)$$

مقدار اطلاع سیگنال دریافتی، $\mathbf{y}$، به عنوان کران بالا دارای مقداری است که متناظر با وضعیتی است که در آن $\mathbf{y}$ دارای ویژگیهای نوفه گاوسی سفید است:

$$H(Y) \le W\log\{2\pi e(P_x + P_n)\}. \qquad (۱۷.۶)$$

در این صورت مقدار اطلاع ارسال شده برابر است با

$$I(X;Y) \le W\log\{2\pi e(P_x + P_n)\} - W\log\{2\pi e P_{H_n}\}, \qquad (۱۸.۶)$$

بنابراین کران بالا برای ظرفیت به صورت زیر می‌باشد

$$C \le W\log\left\{\frac{P_x + P_n}{P_{H_n}}\right\}. \qquad (۱۹.۶)$$

کران پایین را می‌توان با در نظر گرفتن نرخ به دست آورد، اگر سیگنال ورودی را با کدگذاری مناسب به یک نوفه سفید با توان p_x تبدیل کنیم و نتیجهٔ بخش (۵.۵) را که توان اطلاع مجموع دو سیگنال بزرگتر یا برابر با مجموع توانهای اطلاع تکی است به کار بریم. داریم

$$P_x + P_{H_n} = P_{H_x} + P_{H_n} \le P_{H_y}. \qquad (۲۰.۶)$$

با به کار بردن کران بالا برای $H(Y)$ به دست می‌آوریم

$$H(Y) = W\log\{2\pi e P_{H_y}\} \geq W\log\{2\pi e(P_x + P_{H_n})\} \qquad (۲۱.۶)$$

و از این رو داریم

$$I(X;Y) \geq W\log\{2\pi e(P_x + P_{H_n})\} - W\log\{2\pi e P_{H_n}\},$$

بنابراین

$$C \geq W\log\left\{\frac{P_x + P_{H_n}}{P_{H_n}}\right\}. \qquad (۲۲.۶)$$

سرانجام، با ترکیب این دو نتیجه ظرفیت کانال را به صورت نابرابری داده شده در زیر به دست می‌آوریم:

$$W\log\left\{\frac{P_x + P_{H_n}}{P_{H_n}}\right\} \leq C \leq W\log\left\{\frac{P_x + P_n}{P_{H_n}}\right\}. \qquad (۲۳.۶)$$

در حالتی که در آن نوفه گاوسی است داریم $P_{H_n} = P_n$ و نابرابری داده شده در بالا به عبارت مذکور قبلی برای ظرفیت کانال تبدیل می‌شود.

مثال ۲.۶

سیگنال ورودی یک کانال نوفه‌ای دارای توان متوسط $P_x = \sigma_x^2 = 4$ است. نوفه دارای توزیع یکنواخت روی $[-\alpha,\alpha]$ است. بنابراین مقدار اطلاع نوفه برابر می‌شود با

$$H(\mathrm{N}) = \log 2\alpha \text{ بیت / نمونه} = 2W\log 2\alpha \text{ بیت / ثانیه}$$

برای توان اطلاع و توان متوسط نوفه، داریم

$$P_{H_n} = \frac{1}{2\pi e}2^{2H(\mathrm{N})} = \frac{2\alpha^2}{\pi e},$$

$$P_n = \int_{-\alpha}^{\alpha} n^2 \frac{1}{2\alpha}dn = \frac{1}{3}\alpha^2.$$

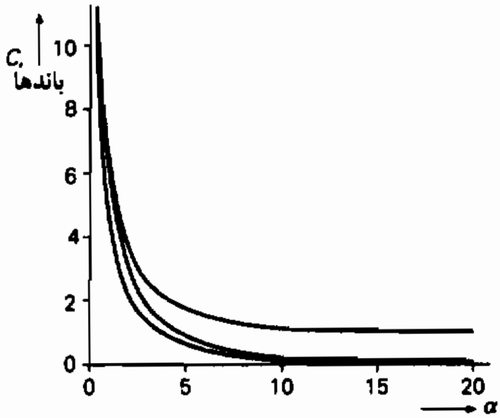
با جانشین کردن توان در معادلهٔ (۲۳.۶) نتیجه می‌شود

$$W\log(1 + \frac{2\pi e}{\alpha^2}) \leq C \leq W\log(\frac{1}{6}\pi e + \frac{2\pi e}{\alpha^2}).$$

در حالتی که در آن نوفه گاوسی باشد با همان توان متوسط $\frac{1}{3}\alpha^2$، ظرفیت برابر است با (با معادلهٔ (۱۵.۶) مقایسه کنید)

$$C = W\log(1+\frac{P_x}{P_n}) = W\log(1+\frac{12}{\alpha^2}).$$

شکل (۳.۶) را ببینید.



شکل ۳.۶- ظرفیت نوفه سفید گاوسی و کرانها برای نوفه سفید غیرگاوسی

۴.۶ قضیهٔ کدگذاری کانال

برای کانال ارتباطی پیوسته نیز حالتی وجود دارد که می‌توان اطلاع را با خطای کوچک دلخواه ε منتقل کرد به شرط آن که مقدار اطلاع ورودی کانال کوچکتر از ظرفیت کانال باشد. از این رو مشابه قضیهٔ کدگذاری کانال برای حالت گسسته است که در فصل چهارم ارائه شد.

اثبات خلاصه شده‌ای از این قضیهٔ کدگذاری کانال را با اثبات این که کدهایی وجود دارند که با آنها ممکن است از ظرفیت کانال با خطای کوچک دلخواه ε حداکثر استفاده را برد ارائه خواهیم کرد.

قضیهٔ ۱.۶ (قضیهٔ کدگذاری شانون برای کانالهای پیوسته)

انتقال مقدار اطلاع $H(X)$ از طریق کانال گاوسی سفید پیوسته با ظرفیت C با احتمال خطای کوچک دلخواه ε (یا ابهام) در حالتی که در آن $H(X) \le C$ امکان‌پذیر است.

برهان

منبع اطلاع پیوسته یک سیگنال تصادفی تولید می‌کند که همواره در این بخش آن

را روی یک مدّت زمان معیّن T در نظر می‌گیریم. با این فرض که منبع اطلاع ارگودیک است، ویژگیهای تصادفی منبع وضع یکسانی در تشخیص سیگنال تصادفی با مدّت T دارند.

اکنون $M(T)$ سیگنال مختلف هر یک با مدّت T را در نظر می‌گیریم و فرض می‌کنیم که همهٔ این سیگنالها دارای یک احتمال هستند. در این صورت به سادگی دیده می‌شود که مقدار اطلاع منبعی را که این سیگنالها را تولید می‌کند می‌توان به صورت زیر نوشت

$$H(X)=\lim_{T\to\infty}\frac{1}{T}\log M(T)\ \text{ثانیه / بیت}. \qquad (۲۴.۶)$$

همان‌طور که قبلاً در این فصل برای ظرفیت کانال در حالت نوفهٔ گاوسی سفید به دست آوردیم؛ معادلهٔ (۱۵.۶) را ببینید. داریم

$$C=W\log\left\{1+\frac{P_x}{P_n}\right\}\ \text{ثانیه / بیت}.$$

اکنون باید ثابت کنیم که ε احتمال خطا می‌تواند صفر شود اگر $H(X)<C$، یا به عبارت دیگر اگر

$$\lim_{T\to\infty}\frac{1}{T}\log M(T)<W\log\left\{1+\frac{P_x}{P_n}\right\}. \qquad (۲۵.۶)$$

برای اثبات قضیه از روشی که در آن سیگنالها به صورت بردارهایی که در یک فضای بعد بالا نمایش داده می‌شوند استفاده خواهیم کرد. بنابر قضیهٔ نمونه‌گیری $2WT$ نمونه نیاز داریم برای این که یک سیگنال پیوسته با مدّت T و با پهنای باند W هرتز را نمایش دهیم. اکنون با معرّفی یک فضای $2WT$- بعدی نمایش سیگنالها به صورت بردارهایی که در آن مقادیر عناصر متناظر با نمونه‌ها هستند و نوک آنها نقاطی در فضای $2WT$- بعدی را تعریف می‌کنند.

از این رو یک سیگنال موزیک با پهنای باند ۱۵ کیلوهرتز و مدّت ۶۰ دقیقه را می‌توان به صورت نقطه‌ای در فضای 10^8-بعدی نمایش داد. سیگنالهای ارسال شده یا دریافت شده همچنین سیگنالهای نوفه را می‌توان به عنوان بردارها یا نقاط به این طریق نمایش داد.

$2WT$ نمونه را با x_i، $i=1,\ldots,2WT$، نمایش می‌دهیم. d طول سیگنال برداری برابر خواهد بود با

$$d = \left(\sum_{i=1}^{2WT} x_i^2\right)^{\frac{1}{2}}. \qquad (۲۶.۶)$$

اگر سیگنالها دارای متوسط مقدار صفر باشند با توجّه بـه ایـن کـه تـوان متوسـط (واریانس) به صورت زیر داده شده است:

$$P = \frac{1}{2WT}\sum_{i=1}^{2WT} x_i^2, \qquad (۲۷.۶)$$

به دست می‌آوریم

$$d = \sqrt{2WTP}. \qquad (۲۸.۶)$$

از این رو تمام سیگنالها با توان P روی یک ابرکره‌ای بـا شـعاع $d = \sqrt{2WTP}$ در فضای $2WT$-بعدی قرار دارند. در این جا می‌توان یک ابرکره را تعمیم $2WT$-بعدی یـک کرهٔ سه-بعدی نرمال در نظر گرفت. به طور مشابه همچنین از ابرحجم علاوه بر حجم کره سخن می گوییم.

حجم یک ابرکره $2WT$-بعدی متناسب با شعاع d به توان $2WT$ است:

$$V_{2WT} = \alpha_{2WT} d^{2WT}, \qquad (۲۹.۶)$$

که در آن α_{2WT} با

$$\alpha_{2WT} = \frac{\pi^{WT}}{\Gamma(WT+1)}, \qquad (۳۰.۶)$$

داده شده است که در آن $\Gamma(.)$ تابع گاماست.

اکنون فرض می کنیم سیگنالی داریم با تـوان متوسـط P_x، کـه نوفـه‌ای بـا تـوان متوسط P_n بر آن تأثیر گذاشته است. هر سیگنال ورودی را می‌توان به عنوان نقطـه‌ای در فضای $2WT$-بعدی، یعنی جایی در ابرکره با شعاع $d = \sqrt{2WTP_x}$ نشان داد.

سیگنالهای خروجی ممکن نقاطی در ابرکره با شعاع $d = \sqrt{2WT(P_x + P_n)}$ هستند.

موقعیت دقیق نقطه یک سیگنال خروجی که شامل سیگنال ورودی به اضافهٔ نوفـه است معلوم نیست، با وجود این، باید جایی در کره‌ای با شـعاع $d = \sqrt{2WTP_n}$ بـه مرکـز نقطه‌ای که سیگنال ارسال شده را نمایش می‌دهد باشد. در شکل (۴.۶) این کره بـا یـک دایرهٔ سایه‌دار نشان داده شده است.

بدیهی است اگر ابرکره نسبت به نوفه دارای هیچ هم‌پوشی نباشد، هیچ تعویضی بین سیگنال ارسال شده و دریافت شده نخواهد بود. در این حالت ارسال اطلاع می‌تواند بـدون

خطا باشد. فرض کنید $M(T)$ تعداد سیگنالهای ارسال شده باشد به قسمی که ابرکره‌های کوچک با نوفه هم‌پوشی ندارند. حجم ابرکره سیگنالهای خروجی حداقل $M(T)$ برابر حجم یکی از ابرکره‌های نوفه خواهد بود. از این رو

$$\alpha_{2WT}\left(\sqrt{2WT(P_x+P_n)}\right)^{2WT} \geq M(T)\alpha_{2WT}\left(\sqrt{2WTP_n}\right)^{2WT},$$

و

$$M(T) \leq \left(\sqrt{\frac{P_x+P_n}{P_n}}\right)^{2WT} = \left(1+\frac{P_x}{P_n}\right)^{WT}. \qquad (۳۱.۶)$$

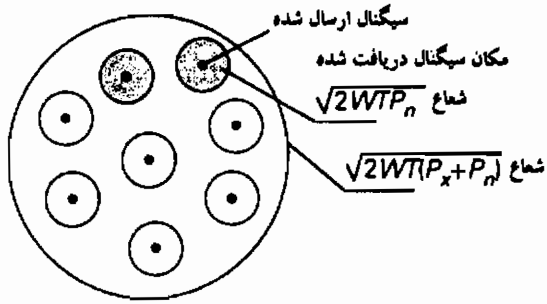
نسبت $\frac{P_x}{P_n}$ مجدداً نسبت سیگنال به نوفه مشهور است. از این نتیجه می‌شود که

$$\lim_{T\to\infty} \frac{\log M(T)}{T} \leq W\log\left(1+\frac{P_x}{P_n}\right) \qquad (۳۲.۶)$$

از این رو مقدار اطلاع کمتر از ظرفیت است.

می‌توان این موضوع را که ممکن است یک نرخ ارسال اطلاع کاملاً نزدیک به ظرفیت با احتمال خطای ε که به صفر میل می‌کند به دست آورد به صورت زیر مشاهده کرد.

تعداد معیّن $M(T)$ نقطه از این فضا منسوب به سیگنال را، بدون توجّه به فاصله‌های متقابل آنها برای اجتناب از هم‌پوشی نواحی، ثابت نگه می‌داریم. انتخاب خاص $M(T)$ نقطهٔ کدگذاری خاصی را برای سیگنالهایی که باید ارسال شوند به وجود می‌آورد. اگر نقطه‌ای در داخل ابرکره نوفهٔ نقطه دیگری باشد خطاها رخ می‌دهند؛ در این صورت امکان دارد به طور نادرست تشخیص داده شود. p احتمال این که این نقطه در داخل یک ابرکره نوفه باشد برابر با نسبت حجمش به حجم ابرکرهٔ بیرونی است:



شکل ۶.۴- ابرکره $2WT$-بعدی با سیگنالهای (تغییر شکل‌یافته)

$$p = \frac{\alpha_{۲WT}\{\sqrt{۲WTP_n}\}^{۲WT}}{\alpha_{۲WT}\{\sqrt{۲WT(P_x + P_n)}\}^{۲WT}}. \qquad (۳۳.۶)$$

احتمال این‌که یک سیگنال در داخل ابرکره نوفه نباشد برابر $p - ۱$ است. برای این که دریافت اطلاع بدون خطا باشد باید $M(T) - ۱$ سیگنال خارج از این ابرکره نوفه باشد. احتمال این برابر $(۱-p)^{M(T)-۱}$ است. این احتمال باید به یک نزدیک شود، از این رو

$$(۱-p)^{M(T)-۱} > ۱-\varepsilon, \qquad (۳۴.۶)$$

که در آن ε یک عدد کوچک دلخواه است.

با بسط سمت چپ به صورت یک سری و قطع آن بعد از جملهٔ دوم، این نابرابری یقیناً برقرار خواهد شد اگر نابرابری زیر برقرار باشد:

$$۱-(M(T)-۱)p > ۱-\varepsilon \quad \Rightarrow \quad (M(T)-۱)p < \varepsilon$$
$$\Rightarrow \quad M(T)-۱ < \frac{\varepsilon}{p} \quad \Rightarrow \quad M(T)-۱ < \varepsilon\left\{\frac{P_x + P_n}{P_n}\right\}^{WT}. \qquad (۳۵.۶)$$

یک شرط قویتر عبارت است از

$$M(T) < \varepsilon\left\{\frac{P_x + P_n}{P_n}\right\}^{WT}. \qquad (۳۶.۶)$$

از این شرط نتیجه می‌شود

$$\lim_{T\to\infty}\frac{۱}{T}\log M(T) < \lim_{T\to\infty}\frac{۱}{T}\log\left[\varepsilon\left\{\frac{P_x + P_n}{P_n}\right\}^{WT}\right]$$
$$= W\log\left(۱+\frac{P_x}{P_n}\right) + \lim_{T\to\infty}\frac{\log\varepsilon}{T}. \qquad (۳۷.۶)$$

حدّ باقی‌مانده در سمت راست با افزایش T به صفر نزدیک می‌شود. بنابراین می‌توان ε را به دلخواه کوچک ساخت به شرط آن که نابرابری $H(X) < C$ برقرار باشد و به شرط آن‌که مدّت T سیگنال را بزرگ انتخاب کنیم. ■

از این رو، کدهایی وجود دارند که ارسال با نرخی به دلخواه نزدیک با ظرفیت کانال با نرخ خطای کوچک دلخواه ε را اجازه می‌دهد. این دیدهٔ اساسی پشتیبان قضیهٔ کدگذاری کانال شانون را تشکیل می‌دهد.

۵.۶ ظرفیت یک کانال گاوسی باحافظه

در بخشهای گذشته ظرفیت برای حالتی که در آن نوفه سفید است تعیین شد، یعنی حالتی که دارای طیف چگالی توان مسطح و چگالی احتمال گاوسی است.

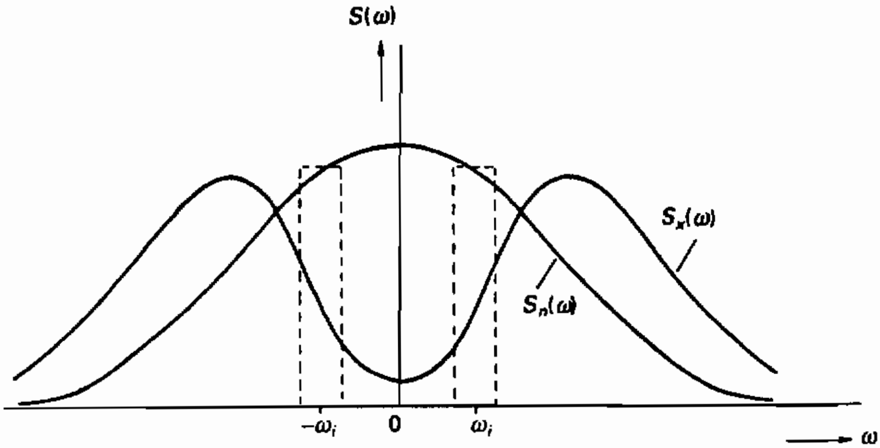
نوفه در عمل غالباً طیف مسطح نخواهد داشت در نتیجه نمونه‌ها به طور آماری مستقل نیستند. برای تعیین ظرفیت کانال در چنین حالتی (کانال باحافظه) باید طیف سیگنالهای ورودی تصادفی و نوفه $\mathbf{x}(t)$ و $\mathbf{n}(t)$ را مورد توجّه قرار داد.

فرض می‌کنیم که سیگنالهای ورودی و نوفه هر دو با توان P و پهنای باند W محدود و دارای چگالی گاوسی باشند. بنابراین رابطهٔ بین توان و توان چگالی طیف برابر است با (همچنین با بخش (۲.۵) مقایسه کنید)

$$P_x = \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} S_x(\omega)\,d\omega, \qquad (۳۸.۶)$$

$$P_n = \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} S_n(\omega)\,d\omega. \qquad (۳۹.۶)$$

یک مثال از دوتا از چنین طیفی در شکل (۵.۶) داده شده است.



شکل ۵.۶- مثال طیف نوفه و سیگنال ورودی

سپس فرض می‌کنیم طیف به ناحیه‌هایی با پهنای $\Delta\omega$ تقسیم شده باشد که آن قدر کوچکند که می‌توان طیف را در هر زیر ناحیه ثابت در نظر گرفت، یعنی $S(\omega_i)$، و سیگنالهای خروجی این صافیهای با پهنای باند (خیلی) کوچک تخیّلی کم و بیش

ناهمبسته‌اند. بنابراین توان متوسط در هر ناحیه $\Delta\omega$ برابر است با

$$P_{x_i} = \frac{1}{2\pi} S_x(\omega_i)\Delta\omega,$$

$$P_{n_i} = \frac{1}{2\pi} S_n(\omega_i)\Delta\omega.$$

فرض کنید تعداد زیرناحیه‌ها N باشد. در این صورت $N = \frac{4\pi W}{\Delta\omega}$ برقرار می‌باشد و از این رو $W = N\omega_i$ که در آن $\omega_i = \frac{\Delta\omega}{4\pi}$. جانشین‌سازی در عبارات مربوط به ظرفیت کانال را برای ظرفیت کانال در یک زیرناحیه به دست می‌آوریم

$$C_i = \omega_i \log\left\{1 + \frac{P_{x_i}}{P_{n_i}}\right\} = \frac{\Delta\omega}{4\pi}\log\left\{1 + \frac{S_x(\omega_i)}{S_n(\omega_i)}\right\}. \qquad (۶.۴۰)$$

ظرفیت برای طیف کامل بین $-2\pi W$ و $2\pi W$ برابر است با

$$C = \sum_{i=1}^{N} C_i,$$

یعنی

$$C = \frac{\Delta\omega}{4\pi}\sum_{i=1}^{N}\log\left\{1 + \frac{S_x(\omega_i)}{S_n(\omega_i)}\right\}. \qquad (۶.۴۱)$$

سپس با حدگیری وقتی $\Delta\omega \to 0$ نتیجه می‌شود که

$$C = \frac{1}{4\pi}\int_{-2\pi W}^{2\pi W}\log\left\{1 + \frac{S_x(\omega)}{S_n(\omega)}\right\}d\omega \text{ بیت / ثانیه}. \qquad (۶.۴۲)$$

اگر طیف زوج باشد، آن‌گاه $S_x(\omega) = S_x(-\omega)$ بنابراین می‌توان نوشت:

$$C = \frac{1}{2\pi}\int_{0}^{2\pi W}\log\left\{1 + \frac{S_x(\omega)}{S_n(\omega)}\right\}d\omega \text{ بیت / ثانیه}. \qquad (۶.۴۳)$$

برای توضیح حالتی را بررسی می‌کنیم که هر دو طیف ثابت هستند. در این صورت داریم

$$P_x = 2WS_x(\omega), \qquad (۶.۴۴)$$

و

$$P_n = 2WS_n(\omega), \qquad (۶.۴۵)$$

که ظرفیت کانال به دست آمده تبدیل می‌شود به

$$C = W \log\left\{1 + \frac{P_x}{P_n}\right\} \text{ثانیه / بیت}, \qquad (۴۶.۶)$$

که دقیقاً عبارتی است که قبلاً برای ظرفیت کانال با نوفهٔ سفید به دست آوردیم.

از این رو معلوم می‌شود که ظرفیت کانال به طیفهای سیگنال و نوفه هر دو بستگی دارد. معمولاً طیف نوفه معلوم است، ولی هنوز هم می‌توان طیف سیگنال را انتخاب کرد. سؤالی که پیش می‌آید این است که بهترین انتخاب برای $S_x(\omega)$ چیست، یا این که ماکسیمم مقدار C برای طیف نوفه داده شده $S_n(\omega)$ چیست.

اکنون عبارت ظرفیت کانال به صورت زیر نوشته می‌شود

$$C = \frac{1}{2\pi}\int_0^{2\pi W} \log[S_x(\omega) + S_n(\omega)]d\omega - \frac{1}{2\pi}\int_0^{2\pi W} \log S_n(\omega)d\omega. \qquad (۴۷.۶)$$

انتگرال دوم برای مقدار معلوم $S_n(\omega)$ ثابت است، بنابراین تنها انتگرال اوّل را می‌توان ماکسیمم کرد. چون توان محدود است، یعنی

$$P_x + P_n = \frac{1}{\pi}\int_0^{2\pi W} (S_x(\omega) + S_n(\omega))d\omega, \qquad (۴۸.۶)$$

مسألهٔ ماکسیممی که باید حل شود در واقع مسألهٔ زیر است. تعیین ماکسیمم

$$I = \int_0^{2\pi W} \log[f(\omega)]d\omega,$$

همراه با قید

$$\int_0^{2\pi W} f(\omega)d\omega = \pi(P_x + P_n).$$

که در آن $f(\omega) = S_x(\omega) + S_n(\omega)$.

از روش لاگرانژ نتیجه می‌شود که

$$\frac{d\log[f(\omega)]}{df(\omega)} - \lambda\frac{df(\omega)}{df(\omega)} = 0,$$

یا

$$\frac{\log e}{f(\omega)} - \lambda = 0,$$

در نتیجه

$$f(\omega) = \frac{\log e}{\lambda} = \text{ثابت}.$$

جانشین‌سازی در قید نشان می‌دهد

$$f(\omega)=\frac{(P_x+P_n)}{۲W},$$

بنابراین

$$I=۲\pi W\log\left\{\frac{(P_x+P_n)}{۲\pi W}\right\}.$$

پس ظرفیت ماکسیمم است اگر

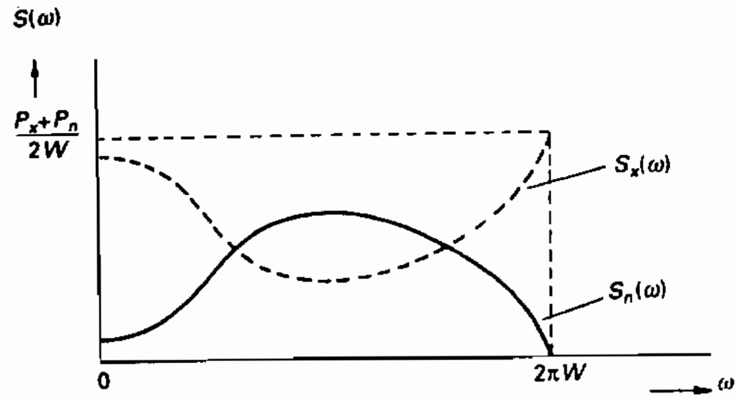
$$S_x(\omega)+S_n(\omega)=\frac{P_x+P_n}{۲W}=\text{ثابت}. \tag{۴۹.۶}$$

در آن حالت ظرفیت کانال برابر است با

$$C=W\log\left\{\frac{P_x+P_n}{۲W}\right\}-\frac{۱}{۲\pi}\int_{۰}^{۲\pi W}\log S_n(\omega)d\omega. \tag{۵۰.۶}$$

مثالی از این در شکل ۶.۶ داده شده است.

همان‌طور که دیده می‌شود $S_x(\omega)$ را باید به طریقی انتخاب کرد ($S_x(\omega)=$ ثابت $-S_n(\omega)$) که چگالی طیفی برای مقادیری از ω بزرگ باشد که در آن نوفه چگالی توان کوچک دارد. نسبت سیگنال به نوفه برای این مقدار ω در حدّ ممکن بزرگ است. به علاوه نتیجه می‌شود که نوفه با چگالی طیف توان مسطح نامناسبترین حالت است، زیرا در این صورت جملهٔ دوم در عبارت ماکسیمم می‌شود که موجب می‌نیمم شدن C می‌شود.



شکل ۶.۶

مثال ۳.۶

طیفهای سیگنالهای تصادفی $x(t)$ و $n(t)$ به ترتیب عبارتند از:

$$S_x(\omega)=\begin{cases} ۴ & , \quad ۰<|\omega|<\pi W, \\ ۸ & , \quad \pi W<|\omega|<۲\pi W, \\ ۰ & , \quad \text{سایر نقاط}. \end{cases}$$

$$S_n(\omega)=\begin{cases} ۱ & , \quad ۰<|\omega|<\frac{۲}{۳}\pi W, \\ ۲ & , \quad \frac{۲}{۳}\pi W<|\omega|<\frac{۴}{۳}\pi W, \\ ۴ & , \quad \frac{۴}{۳}\pi W<|\omega|<۲\pi W, \\ ۰ & , \quad \text{سایر نقاط}. \end{cases}$$

ظرفیت کانال برابر می‌شود با

$$C=\frac{۱}{۲\pi}\int_{۰}^{۲\pi W}\log\left\{۱+\frac{S_x(\omega)}{S_n(\omega)}\right\}d\omega$$

$$=\frac{۱}{۲\pi}\left\{\frac{۲}{۳}\pi W\log(۱+\frac{۴}{۱})+\frac{۱}{۳}\pi W\log(۱+\frac{۴}{۲})+\frac{۱}{۳}\pi W\log(۱+\frac{۸}{۲})+\frac{۲}{۳}\pi W\log(۱+\frac{۸}{۴})\right\}$$

$$=\frac{۱}{۲}W\log ۱۵=۱{,}۴۵W.$$

توان متوسط برابر است با

$$P_x=\frac{۱}{\pi}\int_{۰}^{۲\pi W}S_x(\omega)d(\omega)=\frac{۱}{\pi}(۴\pi W+۸\pi W)=۱۲W,$$

$$P_n=\frac{۱}{\pi}\int_{۰}^{۲\pi W}S_n(\omega)\,d\omega=\frac{۱}{\pi}(\frac{۲}{۳}\pi W+\frac{۴}{۳}\pi W+\frac{۸}{۳}\pi W)=\frac{۱۴}{۳}W.$$

از این رو نسبت سیگنال به نوفه برابر است با $\frac{۱۸}{۷}=\frac{۱۲W}{(\frac{۱۴}{۳}W)}$.

ظرفیت ماکسیمم است اگر

$$S_x(\omega)+S_n(\omega)=\frac{P_x+P_n}{۲W}=\frac{\frac{۱۴}{۳}+\frac{۳۶}{۳}}{۲}=\frac{۲۵}{۳}.$$

در این صورت ماکسیمم ظرفیت کانال برابر است با

$$C = W \log\left\{\frac{P_x + P_n}{۲W}\right\} = W \log \frac{۲۵}{۳} = ۲{,}۰۶W. \quad \blacktriangle$$

۶.۶ تمرینها

۱.۶ سیگنال ورودی یک کانال ارتباطی پیوسته با x و سیگنال خروجی با y نمایش داده می‌شود. برای این کانال چگالی احتمال توأم $p(x,y)$ برای هر دو سیگنال در ناحیهٔ G داده شده با $۱>x>۰$ و $۰<y<x$ یکنواخت است و در خارج این ناحیه $p(x,y)=۰$.

(الف) مقدار اطلاع این منبع را محاسبه کنید.
(ب) مقدار اطلاع را در طرف گیرنده پیدا کنید.
(پ) ایهام را محاسبه کنید.
(ت) مقدار اطلاع برای رخ‌داد توأم x و y را به دست آورید.
(ث) مقدار اطلاع ارسال شده را محاسبه کنید.
(ج) رابطهٔ $H(X,Y) \le H(X)+H(Y)$ را برای این کانال ثابت کنید.

۲.۶ یک منبع اطلاع سیگنالی با چگالی احتمال نمایی-منفی از دامنهٔ نوسان را می‌دهد:

$$p(x) = e^{-x} \quad , \quad ۰ < x < \infty.$$

این منبع به کانالی متّصل شده که با نوفه جمعی مستقل با چگالی احتمال یکنواخت میدان نوسان را بین مقادیر دامنهٔ ۰ و α تغییر شکل می‌دهد. چگالی احتمال خارج از این دامنه صفر است؛ مقدار اطلاع توأم را نسبت به ورودی و خروجی محاسبه کنید.

۳.۶ (الف) عبارتی برای ظرفیت کانال یک کانال پیوسته که توسط نوفه دلخواهی تغییر شکل یافته است به دست آورید. این نوفه جمعی و مستقل از $x(t)$ است (که دارای توان کراندار است).

(ب) ظرفیت کانال را به صورت تابعی از نسبت سیگنال به نوفه $\frac{P_x}{P_n}$ برای پهنای باند ثابتی رسم کنید.
(پ) کران بالا را برای ظرفیت کانال اگر پهنای باند بدون محدودیت افزایش یابد و

اگر فرض شده باشد که میانگین توان نوفه بر واحد پهنای باند ثابت و برابر P_n^o است محاسبه کنید.

(ت) ظرفیت کانال را به صورت تابعی از پهنای باند تحت شرط داده شده در (پ) رسم کنید.

(ث) اگر ظرفیت کانال برای کانالی که قبلاً یادآوری شده برابر $۵٫۶ \times ۱۰^۴$ بیت بر ثانیه باشد و بتوان برای سیگنالی که باید ارسال شود بین دو پهنای باند ۷ و ۸ کیلوهرتز انتخاب کرد، در این صورت چه نسبتهایی از سیگنال به نوفه برای هر دو حالت به دست می‌آید؟

(ج) آیا می‌توانید دربارهٔ برتری ممکن یک سیگنال بر سیگنال دیگر، مطلبی بیان کنید؟

۶.۴ یک منبع اطلاع بی‌حافظه سیگنال $x(t)$ را با چگالی احتمال گاوسی $N(0,1)$ می‌دهد. این منبع به کانالی متّصل شده که توسط نوفهٔ جمعی مستقل با چگالی احتمال دامنهٔ نوسان زیر تغییر شکل می‌یابد.

$$p(n) = n^2 \quad , \quad |n| \le a,$$

$$= 0 \quad , \quad |n| > a.$$

چگالی طیفی توان سیگنال و نوفه بین $-W$ و $+W$ محدود شده‌اند. فراوانی نمونه‌گیری به قسمی است که شروط قضیهٔ نمونه‌گیری برقرار می‌باشند.

(الف) مقدار اطلاع در نمونه را محاسبه کنید.

(ب) توان اطلاع در نمونه را پیدا کنید.

(پ) مقدار اطلاع نوفه بر ثانیه را به دست آورید.

(ت) کران بالا برای مقدار اطلاع بر ثانیه در سیگنال دریافت شده $y(t)$ را در خروجی کانال بیابید.

(ث) کران بالا را برای ظرفیت این کانال به دست آورید.

۶.۵ می‌خواهیم یک سیگنال $x(t)$ روی کانال نوفه‌ای باحافظه ارسال کنیم.

(الف) عبارتی برای ظرفیت کانال برحسب طیف سیگنال و نوفه بیان کنید.

(ب) بنابراین یک رابطه بین طیف سیگنال و طیف نوفه که برای آن ظرفیت برای طیف نوفهٔ داده شده $S_n(\omega)$ ماکسیمم است به دست آورید.

(پ) اگر طیف نوفه به صورت زیر باشد

$$S_n(\omega) = \text{N} \quad , \quad 0 < |\omega| < 2\pi W_1,$$
$$= 2\text{N} \quad , \quad 2\pi W_1 < |\omega| < 2\pi W_2,$$
$$= 0 \quad , \quad \text{سایر نقاط}.$$

ماکسیمم ظرفیت کانال را محاسبه کنید، در حالتی که در آن نسبت سیگنال به نوفه برابر ۲ و $W_2 = 2W_1$ داده شده است.

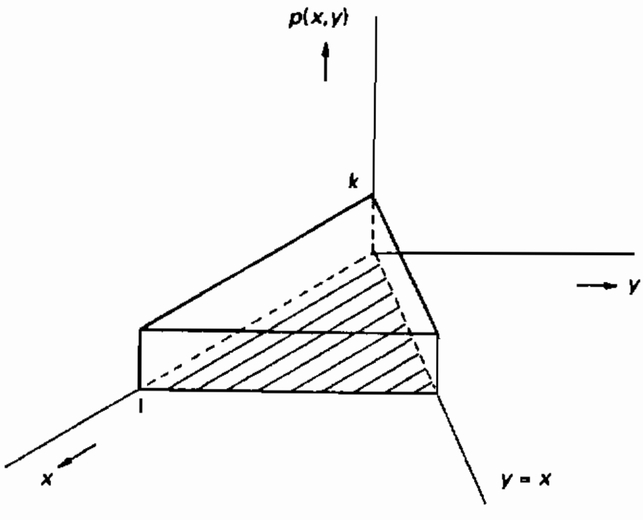
(ت) توان چگالی طیفی سیگنال $S_x(\omega)$ و نوفهٔ $S_n(\omega)$ را در یک شکل رسم کنید. در مورد انتخاب $S_x(\omega)$ چه می‌توانید بگویید؟

۷.۶ جوابها

۱.۶ (الف) شکل (۷.۶) را ببینید.

نخست باید دامنهٔ نوسان چگالی احتمال را تعیین کرد. این کار را می‌توان با استفاده از رابطهٔ زیر انجام داد

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\,dx\,dy = 1.$$



شکل ۷.۶- تعریف دامنهٔ $p(x,y)$ تمرین (۱.۶)

اکنون فرض کنید که $p(x,y)=k$، آن‌گاه

$$\int_0^1\int_0^x k\,dx\,dy = k\int_0^1 dx\int_0^x dy = k\int_0^1 x\,dx = \frac{1}{2}k = 1;$$

بنابراین $p(x,y)=k=2$.

به‌علاوه، باید $p(x)$ را از چگالی احتمال توأم با انتگرال‌گیری روی y محاسبه کرد

$$p(x)=\int_0^x p(x,y)dy = 2y\Big|_0^x = 2x.$$

در این صورت نتیجه می‌شود که

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx$$

$$= -\int_0^1 2x\log 2x\,dx$$

$$= -\frac{1}{2}\int_0^1 2x\log 2x\,d(2x)$$

$$= -\log 2 + \frac{1}{2}\log e = \log\frac{\sqrt{e}}{2} \text{ بیت}.$$

(ب) $H(Y)$ به همین روش محاسبه می‌شود. داریم

$$q(y)=\int_y^1 p(x,y)dx = 2x\Big|_y^1 = 2(1-y),$$

و

$$H(Y) = -\int_0^1 2(1-y)\log 2(1-y)\,dy = -\frac{1}{2}\int_0^1 2z\log 2z\,d(2z).$$

این مانند عبارتی است که برای $H(X)$ در (الف) داریم، بنابراین به سادگی نتیجه می‌شود که

$$H(Y)=H(X)=\log\frac{\sqrt{e}}{2} \text{ بیت}.$$

(پ) ابتدا باید $p(x|y)$ را پیدا کنیم. داریم

$$p(x|y)=\frac{p(x,y)}{q(y)} = \frac{2}{2(1-y)} = \frac{1}{1-y}.$$

اکنون

$$H(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log q(x|y)\,dx\,dy$$
$$= -\int_{۰}^{۱}\int_{۰}^{x} ۲\log\frac{۱}{۱-y}\,dx\,dy = ۲\int_{۰}^{۱}\int_{۰}^{x}\log(۱-y)\,dx\,dy.$$

بعد از انتگرال‌گیری جزء به جزء و جانشین کردن حدود نتیجه می‌شود:

$$H(X|Y) = -\log\sqrt{e}\ \text{بیت}.$$

(ت) این برابر است با

$$H(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x,y)\,dx\,dy = -۲\int_{۰}^{۱}\int_{۰}^{x} dx\,dy = -۱\ \text{بیت}.$$

(ث) می‌توان از رابطهٔ زیر استفاده نمود

$$\mathrm{R} = H(X) - H(X|Y)$$

این نتیجه می‌دهد

$$\mathrm{R} = \log\frac{\sqrt{e}}{۲} + \log\sqrt{e} = \log\frac{e}{۲}\ \text{ثانیه / بیت}.$$

(ج) رابطه عبارت است از

$$H(X,Y) \le H(X) + H(Y)$$

با قرار دادن مقادیر به دست آمده در این رابطه نتیجه می‌شود

$$-۱ \le \log\frac{\sqrt{e}}{۲} + \log\frac{\sqrt{e}}{۲} \quad\Rightarrow\quad -۱ \le \log\frac{e}{۴}.$$

از این رو رابطه به درستی برقرار می‌باشد.

۲.۶ چگالیهای احتمال دامنهٔ نوسان زیر برای سیگنال و نوفه داده شده‌اند:

$$p(x) = e^{-x}\quad,\quad ۰ \le x < \infty,$$
$$p(n) = \frac{۱}{\alpha}\quad,\quad ۰ \le n \le \alpha\quad \left(\text{چون}\ \int_{۰}^{\alpha} p(n)\,dn = ۱\right).$$

برای مقدار اطلاع توأم داریم

$$H(X,Y) = H(X) + H(Y|X) = H(X) + H(\mathrm{N}),$$

چون نوفه از سیگنال x ارسال شده مستقل است. اکنون می‌توان محاسبه کرد که

$$H(X) = -\int_{\circ}^{\infty} e^{-x} \log e^{-x}\, dx = \log e \text{ بیت}$$

و این که

$$H(\mathrm{N}) = -\int_{\circ}^{\alpha} \frac{۱}{\alpha} \log \frac{۱}{\alpha}\, dn = \log \alpha \text{ بیت},$$

سرانجام، داریم

$$H(X, Y) = \log e + \log \alpha = \log \alpha e \text{ بیت}.$$

۳.۶ (الف) ظرفیت این کانال عبارت است از

$$C = W \log\left\{۱ + \frac{P_x}{P_n}\right\}.$$

(ب) این ظرفیت یک تابع لگاریتمی از نسبت سیگنال به نوفه است (شکل (۸.۶) را ملاحظه کنید).

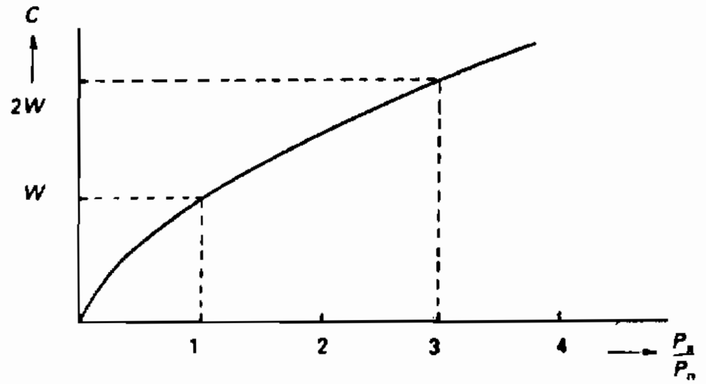
(پ) اکنون توان نوفه برابر است با

$$P_n = W P_n^{\circ},$$

که در آن W پهنای باند نوفه است. در این صورت ظرفیت کانال برابر می‌شود با

$$C = W \log\left\{۱ + \frac{P_x}{W P_n^{\circ}}\right\} = \log e \,.\, \ln\left\{۱ + \frac{P_x}{W P_n^{\circ}}\right\}^{W}$$

$$= \log e \frac{P_x}{P_n^{\circ}} \ln\left\{۱ + \frac{P_x}{W P_n^{\circ}}\right\}^{W P_n^{\circ}/P_x}$$



شکل ۸.۶- ظرفیت به عنوان تابعی از نسبت سیگنال به نوفه

به قسمی که

$$\lim_{W\to\infty} C = \frac{P_x}{P_n^{\circ}}\log e.$$

(ت) رسم نموداری $C = f(W)$ شکل (۹.۶) را می‌دهد.

(ث) $$C = W\log\left\{1+\frac{P_x}{P_n}\right\}$$

بنابراین در یک حالت داریم

$$\frac{P_x}{P_n} = 2^{C/W} - 1 = 2^8 - 1 = 255,$$

و در حالت دیگر

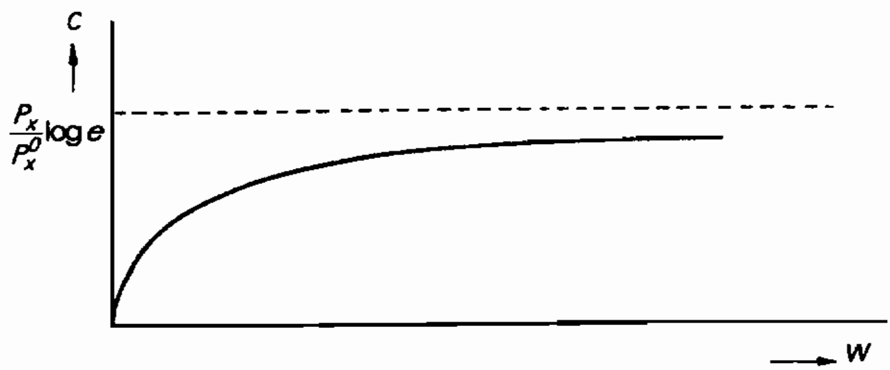
$$\frac{P_x}{P_n} = 2^7 - 1 = 127.$$

(ج) هیچ برتری یک حالت بر حالت دیگر در زمینه‌های نظری اطلاع وجود ندارد، زیرا فقط ظرفیت مهم است. برتری ممکن باید براساس معیار دیگری باشد نظیر این حقیقت که گاهی اوقات به دست آوردن دو برابر نسبت سیگنال به نوفه از ۱۴% پهنای باند بزرگتر مشکلتر است.

۴.۶ (الف) شکل (۶.۱۰) را ببینید.

ممکن است (الف) را با کمک رابطهٔ زیر محاسبه کرد

$$\int_{-a}^{a} p(n)dn = \int_{-a}^{a} n^2\,dn = \frac{1}{3}n^3\Big|_{-a}^{a} = \frac{2}{3}a^3 = 1.$$



شکل ۹.۶- ظرفیت به عنوان تابعی از پهنای باند

از این نتیجه می‌شود $a = \sqrt[3]{\frac{3}{2}}$.

اکنون $H(\mathrm{N})$ را می‌توان به صورت زیر نوشت

$$H(\mathrm{N}) = H(Y|X) = -\int_{-\infty}^{\infty} p(n)\log p(n)\,dn$$

$$= -\int_{-a}^{a} n^2 \log n^2\, dn = -\frac{2}{3}\int_{-a}^{a} \log n\, dn^3$$

$$= -\frac{2}{3}n^3 \log|n|\Big|_{-a}^{a} + \frac{2}{3}\int_{-a}^{a} n^3\, d\log n$$

$$= [-\frac{2}{3}n^3 \log|n| + \frac{2}{3}\log e \frac{1}{3}n^3]\Big|_{-a}^{a}$$

$$= -\frac{4}{3}a^3 \log a + \frac{4}{9}a^3 \log e = \frac{2}{3}\log\frac{2e}{3} = ۰{,}۵۷ \text{ نماد / بیت}.$$

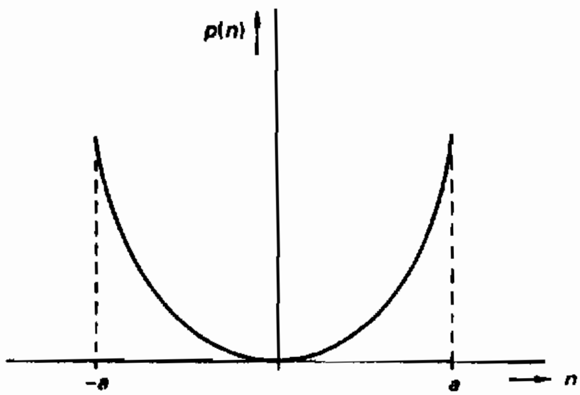
(ب) توان اطلاع نوفه به صورت زیر تعریف می‌شود

$$P_{H_n} = \frac{1}{2\pi e}2^{2H(N)},$$

به قسمی که

$$P_{H_n} = \frac{1}{2\pi e}2^{2\times 0{,}57} = \frac{1}{2\pi e}2^{1{,}14} = 0{,}129.$$

(پ) مقدار اطلاع نوفه بر ثانیه برابر است با



شکل (۶. ۱۰) چگالی احتمال نوفه تمرین (۶.۴)

$$H(\mathrm{N})_{\text{ثانیه}} = ۲WH(\mathrm{N})_{\text{نمونه}} = ۲W \times ۰{,}۵۷ = ۱{,}۱۴W \ \text{ثانیه / بیت}$$

یا

$$H(\mathrm{N})_{\text{ثانیه}} = W\log(۲\pi e P_{H_n}) = ۱{,}۱۴W \ \text{ثانیه / بیت}.$$

(ت) مقدار اطلاع سیگنال دریافت شده $y(t)$ دارای مقدار کران بالای زیر است، یعنی اگر دارای ویژگیهای نوفه مطلوب باشد:

$$H(Y) = W\log ۲\pi e(P_x + P_n),$$

با

$$P_x = \sigma_x^۲ = ۱,$$

و

$$P_n = \sigma_n^۲ = E(\mathbf{n}^۲) = \int_{-a}^{a} n^۲ p(n)\,dn = \int_{-a}^{a} n^۲ n^۲\,dn = \frac{۱}{۵}n^۵\Big|_{-a}^{a} = \frac{۲}{۵}a^۵$$
$$= \frac{۲}{۵}\left(\frac{۳}{۲}\right)^{۵/۳},$$

به طوری که

$$H(Y) = W\log ۲\pi e\left(۱ + \frac{۲}{۵}\left(\frac{۳}{۲}\right)^{۵/۳}\right).$$

(ث) در این صورت مقدار اطلاع ارسال شده برابر است با

$$\mathrm{R} \le H(Y) - H(Y|X) \le H(Y) - H(\mathrm{N})$$
$$\le W\log ۲\pi e(P_x + P_n) - W\log ۲\pi e P_{H_n},$$

به طوری که کران بالا برای ظرفیت به صورت زیر به دست می‌آید

$$C \le W\log\left\{\frac{P_x + P_n}{P_{H_n}}\right\} = W\log\left\{\frac{۱ + \frac{۲}{۵}\left(\frac{۳}{۲}\right)^{۵/۳}}{۰{,}۱۲۹}\right\}.$$

۵.۶ (الف) برای ظرفیت کانال باحافظه به دست می‌آوریم که

$$C = \frac{۱}{۴\pi}\int_{-۲\pi W}^{۲\pi W} \log\left\{۱ + \frac{S_x(\omega)}{S_n(\omega)}\right\} d\omega \ \text{ثانیه / بیت}.$$

طیف سیگنال و نوفه زوج هستند، یعنی نسبت به $\omega = ۰$ متقارن هستند، به طوری که می‌توان نوشت

$$C=\frac{1}{2\pi}\int_0^{2\pi W}\log\left\{1+\frac{S_x(\omega)}{S_n(\omega)}\right\}d\omega \text{ ثانیه / بیت}.$$

(ب) عبارت برای ظرفیت کانال را می‌توان به صورت زیر نوشت

$$C=\frac{1}{2\pi}\int_0^{2\pi W}\log[S_x(\omega)+S_n(\omega)]d\omega-\frac{1}{2\pi}\int_0^{2\pi W}\log S_n(\omega)d\omega.$$

چون طیف نوفه داده شده است، ماکسیمم C به معنی ماکسیمم انتگـــرال اوّل است. این رابطه بین $S_n(\omega)$ و $S_x(\omega)$ به صورت زیر در می‌آید:

$$S_x(\omega)+S_n(\omega)=\frac{P_x+P_n}{2W}=\text{ثابت}.$$

(پ) عبارت ماکسیمم ظرفیت کانال با کمک نتیجهٔ (ب) به دســـت می‌آیـد (همچنیــن شکل (۱۱.۶) را ببینید).

$$C=W\log\left\{\frac{P_x+P_n}{2W}\right\}-\frac{1}{2\pi}\int_0^{2\pi W}\log S_n(\omega)d\omega.$$

برای توان نوفه P_n به دست می‌آوریم

$$P_n=\frac{1}{\pi}\int_0^{2\pi W_1}\mathrm{N}d\omega+\frac{1}{\pi}\int_{2\pi W_1}^{2\pi W_2}2\mathrm{N}d\omega=\frac{\mathrm{N}}{\pi}\omega\Big|_0^{2\pi W_1}+\frac{2\mathrm{N}}{\pi}\omega\Big|_{2\pi W_1}^{2\pi W_2}$$

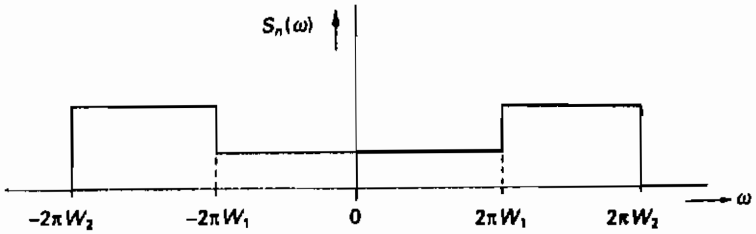
$$=2\mathrm{N}W_1+4\mathrm{N}(W_2-W_1).$$

با استفاده از این امر که $W_2=2W_1$، P_n برابر می‌شود با

$$P_n=\mathrm{N}W_2+2\mathrm{N}W_2=3\mathrm{N}W_2.$$

علاوه بر این نسبت سیگنال به نوفه $\frac{P_x}{P_n}=2$ داده شده است، بنابراین داریم

$$P_x=2P_n=6\mathrm{N}W_2.$$



شکل ۱۱.۶- طیف نوفه تمرین (۵.۶)

با جانشین کردن در فرمول ظرفیت کانال C نتیجه می‌شود

$$C = W_2 \log\left\{\frac{6NW_2 + 3NW_2}{2W_2}\right\} - \frac{1}{2\pi}\int_0^{2\pi W_1} \log N\, d\omega - \frac{1}{2\pi}\int_{2\pi W_1}^{2\pi W_2} \log 2N\, d\omega$$

$$= W_2 \log 4{,}5N - W_1 \log N - \log 2N.[W_2 - W_1]$$

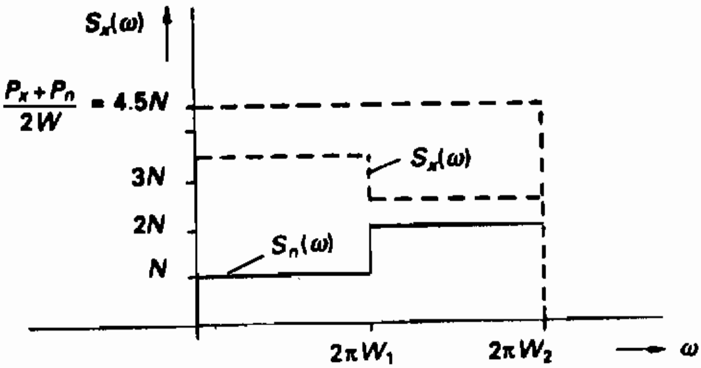
$$= W_2 \log 4{,}5N - \frac{1}{2}W_2 \log N - \frac{1}{2}W_2 \log 2N$$

$$= W_2 \log \frac{4{,}5N}{\sqrt{N}\sqrt{2N}} = W_2 \log \frac{4{,}5}{\sqrt{2}} = W_2 \log 3{,}15$$

$= 1{,}66W_2$ بیت / ثانیه.

(ت) شکل (۶٫۱۲) را ببینید.

طیف چگالی توان $S_x(\omega)$ برای مقادیری از ω که نوفه در آنها یک طیف چگالی توان $S_n(\omega)$ کوچک دارد بزرگ است.



شکل ۶٫۱۲

۷

نظریهٔ نرخ دگرشکلی

۱.۷ تابع نرخ دگرشکلی گسسته

نظریهٔ بررسی‌شده تاکنون دو ویژگی ضروری را نشان می‌دهد. اوّلین ویژگی امکان ارسال بدون خطاست، اگر مقدار اطلاعی که باید ارسال شود کمتر از C ظرفیت کانال باشد، همان‌طور که با قضیهٔ کدگذاری کانال (قضیهٔ (۵.۴)) ثابت شد. برای این که این مطلب درست باشد، پیامها باید بتوانند شامل تعداد بی‌شماری از نمادها باشند. ویژگی دوم این است که فرض می‌کنیم کانالی داده شده و منبع را با این کانال وفق می‌دهیم. از این رو ظرفیت کانال به عنوان ماکسیمم تعداد اطلاع ارسال شده تعریف می‌شود، که در آن ماکسیمم با اتّصال همهٔ منابع ممکن به کانال به دست می‌آید.

این دو ویژگی همواره در عمل مناسب نیستند. در عمل احتمال خطایی وجود خواهد داشت که ممکن است کوچک باشد، ولی البته برابر صفر نیست. برای یک منبع اطلاع پیوسته حتّی بازسازی کاملاً درست درگیرنده امکان‌پذیر خواهد بود. همواره مقداری دگرشکلی وجود خواهد داشت، هر چند کوچک باشد. به علاوه، منبع را اغلب باید همان‌طور که در عمل داده شده نگریست (در نظر گرفتن سیگنالهای یک سخنرانی یا سیگنالهای تلویزیون) و باید کانال با منبع وفق داده شود. این کار با اختیار منبعی با ظرفیت کافی برای دستیابی به یک ارسال خوب انجام می‌شود.

طبیعی است که دگرشکلی را به عنوان چیزی که اجتناب ناپذیر است در نظر بگیریم، این کاملاً امکان‌پذیر است که دگرشکلی عملاً برای برخی حالتها معرّفی شود. با

تراکم (فشردگی) داده‌ها یک دنباله از نمادهای خروجی منبع به دنباله‌ای از نمادهایی از الفبای دوباره تولید شده انتقال داده می‌شود به طریقی که آنتروپی دنبالهٔ جدید کمتر از آنتروپی دنبالهٔ اصلی است، ولی به بهای مقداری دگرشکلی. در واقع باید یک تناظر بین هر دنبالهٔ ورودی و دنبالهٔ خروجی وجود داشته باشد؛ دنبالهٔ ورودی یکسان به دنبالهٔ خروجی یکسان منتهی می‌شود. تراکم داده‌ها یک فرآیند قطعی است. با وجود این، در سطح نمادهای تکی رابطهٔ بین نماد منبع u_j و نماد $\hat{u}_k$ هنگامی که رمزگذاری شده است یک به یک نیست، بلکه متأثر از احتمال انتقال $q(\hat{u}_k|u_j)$ می‌باشد، گرچه در سطح دنباله یا بلوک رابطه قطعی است.

از این فصل ممکن است نتیجه‌گیری کرد که کد داده‌های فشرده به‌خوبی طراحی شده، نمادهای تولید شده کد واژه‌ها باید به قسمی باشد که ماتریس احتمالهای شرطی نمادهای تولید شده معلوم نمادهای منبع، اطلاع متقابل بین واژه‌های منبع و کد واژه‌های مقید به متوسط دگرشکلی را می‌نیمم کند.

نشان ویژه نظریهٔ نرخ دگرشکلی آن است که ترکیبی از منبع و تولید مجدّد را بررسی می‌کند. یک اندازهٔ دگرشکلی برای این ترکیب اجرایی را نشان می‌دهد که برای سیگنالهای تولید شده یا نمادهای تولید شده توسط منبع می‌توان انتظار داشت.

قبل از معرّفی اندازهٔ کمّی برای دگرشکلی، ابتدا منبع اطلاع با جزئیات بیشتری بررسی خواهد شد. این منبع را منبع گسستهٔ مانا فرض می‌کنیم. به علاوه فرض می‌کنیم که منبع بی‌حافظه است بنابراین دارای هیچ ویژگی مارکوفی نمی‌باشد.

فرض شده است که نمادهای منبع از یک الفبای متناهی شامل n نماد مفروض با $\{u_1,\dots,u_j,\dots,u_n\}$ می‌باشد. منبع خروجی برحسب ***الفبای تولید مجدّد*** یا ***الفبای کد واژه*** شامل m نماد $\{\hat{u}_1,\dots,\hat{u}_k,\dots,\hat{u}_m\}$ تولید شده است.

اندازهٔ دگرشکلی که باید تعریف شود می‌بایستی به هر ترکیبی از نماد منبع u_j و نماد تولید مجدّد $\hat{u}_k$ وزن معیّنی را نسبت دهد. بدین منظور تابع نامنفی ***اندازهٔ دگرشکلی نماد*** را به صورت $\rho(j,k)=\rho(u_j,\hat{u}_k)$ تعریف می‌کنیم؛ دگرشکلی مربوط به تولید نماد $\hat{u}_k$ می‌باشد در حالی که u_j نماد منبع تولید شده است. به علاوه فرض می‌کنیم که این دگرشکلی مستقل از وضعیتی (یا زمانی) باشد که در آن نمادها ظاهر می‌شوند.

دگرشکلی توسط ***ماتریس دگرشکلی*** نشان داده می‌شود که دگرشکلی بین نمادهای منبع و مقصد را تشریح می‌کند. برای سیستمی با دو نماد منبع u_1 و u_2 و دو نماد مقصد $\hat{u}_1$ و $\hat{u}_2$، مثلاً می‌توان دگرشکلی را به صورت زیر انتخاب کرد

$$\rho(u_j, \hat{u}_k) = 0 \qquad , \qquad j = k,$$
$$= 1 \qquad , \qquad j \neq k.$$

این نمودار شکل (۱.۷) را می‌دهد.

با این انتخاب ماتریس دگرشکلی، رابطه‌ای با خطاهایی که می‌تواند بین منبع و تولید مجدّد رخ دهد به وجود آمده است. یک تولید مجدّد درست هیچ دگرشکلی به وجود نمی‌آورد و تولید مجدّد نادرست دگرشکلی یک را معرّفی می کند.

یک ماتریس دگرشکلی نباید متقارن باشد از این رو ترکیب u_1 و $\hat{u}_2$ می‌تواند دارای دگرشکلی $\rho(1,2)$ باشد که خیلی بزرگتر از $\rho(2,1)$ است. چون نمادهای منبع دارای احتمال رخ‌داد معیّنی می‌باشند، مرحلهٔ بعد تعیین متوسط دگرشکلی است که اگر یک دنبالهٔ طولانی از نمادها تولید شده باشد رخ می‌دهد. احتمال نماد تولید مجدّد $\hat{u}_k$ اگر u_j ارسال شده باشد با $q(\hat{u}_k|u_j) = q(k|j)$ نشان داده می‌شود. اگر $p(j)$ احتمال رخ‌داد نماد منبع u_j باشد، در این صورت احتمال رخ‌داد توأم u_j و $\hat{u}_k$ به صورت زیر می‌باشد

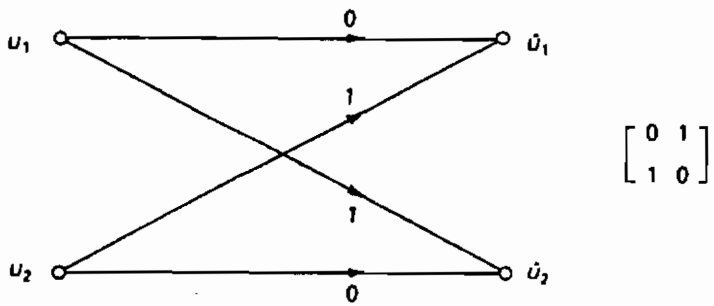
$$p(j,k) = p(j)q(k|j).$$

اکنون ***متوسط دگرشکلی*** $d(Q)$ را با ضرب دگرشکلی هر ترکیب u_j و $\hat{u}_k$ در احتمال رخ‌دادش به دست می‌آوریم

$$d(Q) = \sum_j \sum_k p(j)q(k|j)\rho(j,k). \qquad (۱.۷)$$

اگر فرض شده باشد که دگرشکلی مجاز برابر D است، آن‌گاه باید نابرابری زیر برقرار باشد

$$d(Q) \leq D. \qquad (۲.۷)$$



شکل ۱.۷- مثالی از دگرشکلیها و ماتریس آنها

در عبارت متوسط دگرشکلی $d(Q)$، منبع و از این رو $p(j)$ و دگرشکلی $\rho(j,k)$ معلوم‌اند. احتمالهای انتقال $q(k|j)$ که انتقال بین منبع و تولید مجدّد را نشان می‌دهد ممکن است آزادانه انتخاب شود به قسمی که متوسط دگرشکلی تابعی از این احتمالهای انتقال، یعنی ماتریس Q انتقال یا احتمالهای شرطی باشد. اکنون مجموعهٔ Q_D را که شامل ماتریسهای Q احتمالهای انتقال $q(k|j)$، که برای آن $d(Q) \le D$، معرفّی می‌کنیم.

$$Q_D = \{Q : d(Q) \le D\}. \tag{۳.۷}$$

مثال ۱.۷

برای الفبای منبع و تولید مجدّد به ترتیب $n = ۲$ و $k = ۳$ داده شده‌اند، احتمالهای نمادهای منبع عبارتند از

$$p(u_۱) = p(u_۲) = \frac{۱}{۲}.$$

ماتریس دگرشکلی عبارت است از

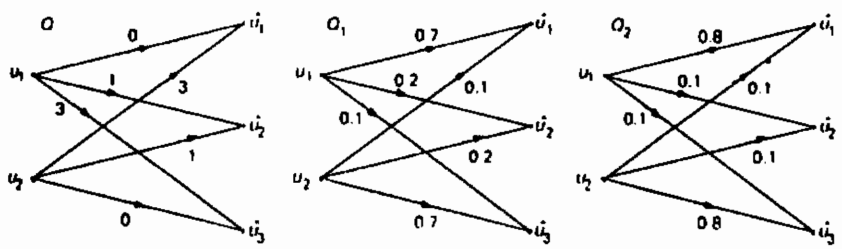
$$Q = \begin{bmatrix} ۰ & ۱ & ۳ \\ ۳ & ۱ & ۰ \end{bmatrix},$$

که نشان می‌دهد اگر نماد منبع $u_۱$ به $\hat{u}_۱$ و $u_۲$ به $\hat{u}_۳$ رمزگذاری شده باشد هیچ دگرشکلی وجود ندارد، در غیر این صورت دگرشکلی وجود دارد.

فرض کنید متوسط مجاز دگرشکلی نماد کوچکتر یا برابر با $D = ۰٫۴۵$ باشد. فرض کنید دو ماتریس دگرشکلی $Q_۱$ و $Q_۲$ به صورت زیر داده شده باشند

$$Q_۱ = \begin{bmatrix} ۰٫۷ & ۰٫۲ & ۰٫۱ \\ ۰٫۱ & ۰٫۲ & ۰٫۷ \end{bmatrix} \quad , \quad Q_۲ = \begin{bmatrix} ۰٫۸ & ۰٫۱ & ۰٫۱ \\ ۰٫۱ & ۰٫۱ & ۰٫۸ \end{bmatrix}.$$

ممکن است برای هر دو ماتریس بررسی شود که آیا آنها متوسط دگرشکلی نمادی که در قید معلوم صدق می‌کند می‌دهند یا خیر.



شکل ۲.۷- دگرشکلیها و احتمال انتقال مثال (۱.۷)

به دست می‌آوریم:

$$d(Q_1)=\sum_{j=1}^{2}\sum_{l=1}^{2}p(j)q(k|j)\rho(j,k)$$

$$=\frac{1}{2}\times 0/7\times 8+\frac{1}{2}\times 0/2\times 1+\frac{1}{2}\times 0/1\times 3+\frac{1}{2}\times 0/1\times 3+\frac{1}{2}\times 0/2\times 1+\frac{1}{2}\times 0/7\times 0$$

$$=\frac{1}{2}.$$

به طور مشابه به دست می‌آوریم

$$d(Q_2)=\frac{2}{5}.$$

واضح است که تنها Q_2 در شرط صدق می‌کند

$$d(Q_2)\le D=0/45.$$

ظاهراً یک دگرشکلی کمتر از D با ماتریس Q_2 امکان‌پذیر است. بنابراین این ماتریس برخلاف ماتریس Q_1 به Q_D متعلّق است.

شکل ماتریس دگرشکلی ممکن می‌سازد که $d(Q)$ را به عنوان احتمال این که در حین انتقال خطایی رخ دهد ببینیم. از این رو D احتمال خطای مجاز برای تولید مجدّد و Q_D مجموعهٔ ماتریسهایی است که متوسط احتمال خطا کوچکتر از یا حداکثر برابر با احتمال خطای مجاز است. ▲

اکنون دگرشکلی را هنگامی که نمادهای تولید مجدّد اندازه‌ای برای کمیت تولید مجدّد منبع به دست آمده‌اند تعریف می‌کنیم. برای این که به تعریفی از تابع نرخ دگرشکلی برسیم اطلاع متقابل بین منبع و مقصد (تولید مجدّد) را در نظر می‌گیریم. در وضعیت موجود داریم:

$$I(U;\hat{U})=H(U)-H(U|\hat{U})$$

$$=H(\hat{U})-H(\hat{U}|U)$$

$$=\sum_{j}\sum_{k}p(j)q(k|j)\log\frac{p(j,k)}{p(j)\cdot q(k)}. \qquad (۴.۷)$$

این اطلاع متقابل اندازه‌ای برای همبستگی بین U و $\hat{U}$ می‌باشد.

از جنبهٔ نظری نرخ دگرشکلی منبع، یعنی $p(j)$، داده شده است در نتیجه اطلاع متقابل به صورت تابعی از ماتریس Q احتمالهای انتقال $q(k|j)$ در نظر گرفته خواهد شد:

$$I(Q)=\sum_j\sum_k p(j)q(k|j)\log\frac{q(k|j)}{q(k)}. \qquad (۵.۷)$$

هر ماتریس Q می‌تواند مقدار متفاوت $I(Q)$ را بدهد. اکنون ماتریسی که برای آن $I(Q)$ مینیمم است جست‌وجو می کنیم و نرخ منبع برای دگرشکلی مجاز D یا به طور ساده ***تابع نرخ دگرشکلی*** $R(D)$ را به صورت زیر تعریف می کنیم:

$$R(D)=\min_{Q\in Q_D} I(Q) \text{ بیت / نماد}. \qquad (۶.۷)$$

این تابع مینیمم مقدار اطلاع متقابل است که با تغییر دادن همهٔ ماتریسهای Q به دست می‌آید. با وجود این، ماتریسها باید در شرطی که دگرشکلی $d(Q)$ای که از این ماتریس ناشی می‌شود کوچکتر از دگرشکلی مجاز D باشد صدق کند، به عبارت دیگر این که $Q\in Q_D$. فرض شده است که منبع و دگرشکلی معلومند.

توجیه این تعریف در چند قضیهٔ جالب قرار دارد که در بخش (۴.۷) ارائه خواهد شد. معلوم می‌شود که تابع نرخ دگرشکلی $R(D)$ را می‌توان به عنوان مینیمم نرخ لازم برای هر کد بلوکی برای داده‌های فشرده در نظر گرفت، به قسمی که متوسط دگرشکلی نماد کوچکتر یا برابر D باشد. در ورای آن نرخ هیچ کد فشرده‌ای با دگرشکلی کوچکتر یا برابر D وجود ندارد. در بخشهای (۲.۷) و (۳.۷) نخست ویژگیهای تابع $R(D)$ به تفصیل مطالعه خواهد شد.

مثال ۲.۷

در مثال (۱.۷) دو ماتریس انتقال Q_1 و Q_2 بررسی شد که تنها برای Q_2، $Q_2\in Q_D$ با $D=۰٫۴۵$ برقرار می‌باشد.

برای $I(Q)$ چنین داریم

$$I(Q)=\sum_{j=1}^{2}\sum_{k=1}^{3}p(j)\,q(k|j)\log\frac{q(k|j)}{q(k)}.$$

چون

$$q(k)=\sum_{j=1}^{2}p(j)q(k|j),$$

به سادگی داریم

$$q(1)=\frac{2}{5}\quad,\quad q(2)=\frac{1}{5}\quad,\quad q(3)=\frac{1}{5}.$$

در نتیجه

$$I(Q_۲)=۲\{\frac{۱}{۲}\times\frac{۷}{۱۰}\log\frac{۷/۱۰}{۲/۵}+\frac{۱}{۲}\times\frac{۲}{۱۰}\log\frac{۲/۱۰}{۱/۵}+\frac{۱}{۲}\times\frac{۱}{۱۰}\log\frac{۱/۱۰}{۲/۵}\}$$
$$=-\frac{۸}{۵}+\frac{۷}{۱۰}\log ۷\approx ۰{,}۳۷.$$

در ادامه بررسی خواهیم کرد که آیا این $I(Q_۲)$ می‌نیمم است به طوری که $R(D)=I(Q_۲)$. ▲

۲.۷ ویژگیهای تابع $R(D)$

در بررسی ویژگیهای تابع $R(D)$ یک منبع اطلاع بی‌حافظه گسسته را در نظر خواهیم گرفت، درست مانند آنچه که قبلاً داشتیم. فرض کنید که منبع دارای n نماد متفاوت است.

در آغاز، باید ببینیم که $R(D)$ به عنوان تابعی از D چگونه عمل می‌کند. برای چنین منبعی تابع نرخ دگرشکلی نوعاً دارای فرمی است که در شکل (۳.۷) نشان داده شده است.

دامنهٔ $R(D)$ عبارت است از

$$\circ\le R(D)\le H(U). \qquad (۷.۷)$$

این مطلب را می‌توان به صورت زیر نشان داد. از

$$\circ\le H(U|\hat{U})\le H(U),$$

و

$$\circ\le H(U)\le\log n$$

و با

$$I(U;\hat{U})=H(U)-H(U|\hat{U})$$

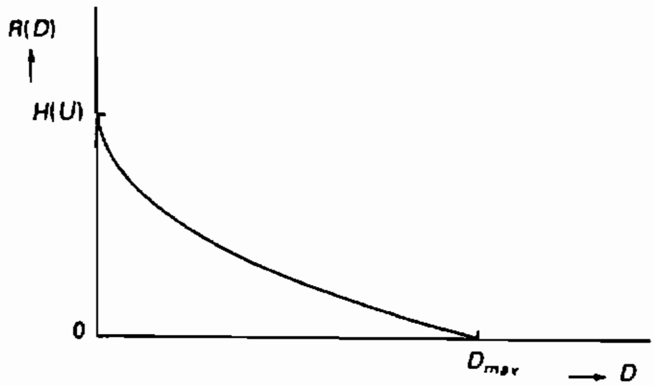
نتیجه می‌شود که

$$\circ\le I(U;\hat{U})\le H(U)\le\log n. \qquad (۸.۷)$$

چون $R(D)$ می‌نیمم $I(U;\hat{U})$ است داریم $R(D)\le I(U;\hat{U})$ که به موجب آن دامنهٔ $R(D)$ به دست می‌آید.

سپس تعیین خواهیم کرد که دگرشکلی مجاز D چه مقدار ممکن را اختیار می‌کند. متوسط دگرشکلی $d(Q)$ به صورت زیر تعریف می‌شود

$$d(Q)=\sum_j\sum_k p(j)q(k|j)\rho(j,k)\le D. \qquad (۹.۷)$$



شکل ۳.۷- تابع نرخ دگرشکلی به صورت تابعی از D

اگر برای هر نماد منبع u_j نماد مقصد $\hat{u}_k$ای جست‌وجو کنیم که $\rho(j,k)$ می‌نیمم باشد کوچکترین متوسط دگرشکلی $D_{\min}$ ممکن به دست می‌آید و سپس این احتمال انتقال را $q(k|j)=1$ قرار داده و بقیهٔ احتمالها را برابر صفر قرار می‌دهیم.

تعریف می‌کنیم

$$\rho(j)=\min_k \rho(j,k). \qquad (۱۰.۷)$$

بنابراین

$$D_{\min}=\sum_{j=1}^{n} p(j)\,\rho(j). \qquad (۱۱.۷)$$

بدون کاستی در کلیّت فرض خواهیم کرد $D_{\min}=0$.

اگر چنین نباشد دگرشکلی $\rho(j,k)$ را به طریقی اصلاح می‌کنیم تا مجدداً $D_{\min}$ برابر صفر شود. این بدین معناست که برای هر نماد u_j بایستی نماد $\hat{u}_k$ با $\rho(j,k)=0$ وجود داشته باشد.

چون برای احتمالهای انتقال داریم

$$q(k|j)=1 \quad , \quad \text{اگر } u_j \text{ و } \hat{u}_k \text{ نظیر یکدیگر باشند (یعنی } q(j,k)=0)$$

$$=0 \quad , \quad \text{برای سایر حالات ,}$$

در این صورت ایهام برابر صفر خواهد شد و $I(Q)=H(U)$. Q دیگر در این جا ظاهر نمی‌شود بنابراین مستقیماً نیز نتیجه می‌شود که

$$R(0)=I(Q)=H(U). \qquad (۱۲.۷)$$

پس برای رسیدن به دگرشکلی صفر باید همهٔ اطلاعات منبع در مقصد تولید شوند.

ماکسیمم متوسط دگرشکلی ممکن، $D_{\max}$ را می‌توان به طریق زیر تعیین کرد. می‌توان نشان داد که تابع نرخ دگرشکلی با افزایش دگرشکلی D یکنواخت نزولی است. در این صورت ماکسیمم D برای مینیمال $R(D)$ رخ می‌دهد، یعنی برای $R(D)=0$. یعنی، مقصد هیچ اطلاعی از منبع دریافت نمی‌کند. در این صورت کوچکترین دگرشکلی ممکن را که هنوز هم امکان‌پذیر است با $D_{\max}$ نشان می‌دهیم؛ این مقدار چنانچه مکرّراً مقدار $\hat{u}_k$ را چنان انتخاب کنیم که برای آن متوسط دگرشکلی بین یک نماد دلخواه منبع و نماد $\hat{u}_k$ مورد سؤال کمترین باشد، به دست می‌آید. هر انتخاب دیگری به متوسط دگرشکلی بزرگتری منتهی می‌شود. اطلاع متقابل $I(Q)$ صفر است اگر $q(k|j)=q(k)$. از یک طرف این از تعریف $I(Q)$ نتیجه می‌شود و از طرف دیگر می‌توان توضیح داد که $q(k|j)=q(k)$ به این معنی است که رخداد u_j هیچ تأثیری بر تولید مجدّد $\hat{u}_k$ ندارد، بنابراین هیچ اطلاعی دوباره تولید نمی‌شود. در این حالت، متوسط دگرشکلی $d(Q)$ برابر می‌شود با

$$d(Q)=\sum_k q(k)\sum_j p(j)\rho(j,k). \qquad (۱۳.۷)$$

می‌نیمم مقدار $d(Q)$ با قرار دادن $q(k)=1$ برای آن مقدار kای که برای آن $\sum_j p(j)\rho(j,k)$ کمترین است به دست می‌آید. این نتیجه می‌دهد

$$D_{\max}=\min_k\sum_j p(j)\rho(j,k).$$

مثال ۳.۷

مثال (۱.۷) را مجدّداً در نظر بگیرید. اکنون برای این که $D_{\min}$ و $D_{\max}$ و مقادیر متناظر با تابع نرخ دگرشکلی را پیدا کنیم.

(الف) $D_{\min}=0$ پیدا می‌شود اگر

$$Q=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}$$

که متناظر است با $R(0)=I(Q)=H(U)=1$.

(ب)
$$D_{\max}=\min_k\sum_{j=1}^{2}p(j)\rho(j,k)$$

$$=\frac{1}{2}\min_k(\rho(1,k)+\rho(2,k))$$

$$= \frac{۱}{۲}\min\{۰+۳, ۱+۱, ۳+۰\} = ۱.$$

مقدار متناظر $R(D)$ در

$$R(D_{\max}) = R(۱) = ۰$$

صدق می‌کند. ماتریس انتقال متناظر عبارت است از

$$Q = \begin{bmatrix} ۰ & ۱ & ۰ \\ ۰ & ۱ & ۰ \end{bmatrix}.$$

▲

تعیین تابع نرخ دگرشکلی را می‌توان به صورت مسأله‌ای از حساب جامعه و فاضلهٔ متغیّرها در نظر گرفت. این را به صورت زیر فرمول‌بندی می‌کنیم: می‌نیمم اطلاع متقابل

$$I(Q) = \sum_j \sum_k p(j) q(k|j) \log \frac{q(k|j)}{q(k)}, \qquad (۱۴.۷)$$

به عنوان تابعی از $q(k|j)$ با قیود زیر:

$$q(k|j) \geq ۰ \qquad \text{(الف)}$$

$$\sum_k q(k|j) = ۱, \qquad (۱۵.۷) \quad \text{(ب)}$$

$$\sum_j \sum_k p(j) q(k|j) \rho(j,k) = D. \qquad (۱۶.۷) \quad \text{(پ)}$$

اگر از قید (الف) صرف‌نظر کنیم با مشتق‌گیری از تابع نسبت به $q(k,j)$ و با قرار دادن نتیجه برابر صفر راه حلی امکان‌پذیر است. مشتق در این جا داده نشده است، ولی عبارت زیر برای احتمالهای انتقال (روش لاگرانژ) به دست می‌آید:

$$q(k|j) = \frac{q(k) e^{s\rho(j,k)}}{\sum_k q(k) e^{s\rho(j,k)}}. \qquad (۱۷.۷)$$

با معرّفی

$$\lambda(j) = \frac{۱}{\sum_k q(k) e^{s\rho(j,k)}}, \qquad (۱۸.۷)$$

نتیجه می‌شود

$$q(k|j) = \lambda(j) q(k) e^{s\rho(j,k)}. \qquad (۱۹.۷)$$

که یک مجموعه‌ای از معادلات برای هر j و k می‌دهد، که در آن $q(k|j)$ برحسب $q(k)$

بیان شده است. بنابراین اکنون تعیین احتمال $q(k)$ باقی می‌ماند. به طور کلّی،

$$q(k)=\sum_j p(j)q(k|j). \tag{۲۰.۷}$$

چنانچه آن را بر $q(k)$ تقسیم کنیم نتیجه می‌دهد

$$\sum_j \lambda(j)\,p(j)\,e^{s\rho(j,k)}=1, \tag{۲۱.۷}$$

یا

$$\sum_j \frac{p(j)\,e^{s\rho(j,k)}}{\sum_k q(k)\,e^{s\rho(j,k)}}=1. \tag{۲۲.۷}$$

اگر قید (الف) برقرار باشد هنوز هم باید ثابت شود. در این روش n معادله برای n احتمال $q(k)$ به دست آمده است، که اکنون ممکن است حل کرد. پس از این ماتریس انتقال Q را می‌توان تعیین کرد. با معادلاتی که این چنین به دست آمده‌اند اکنون ممکن است عبارتی برای $R(D)$ و (D) به دست آورد. جانشین کردن $q(k|j)$ در $d(Q)$ نتیجه می‌دهد

$$D=\sum_j\sum_k \lambda(j)\,p(j)q(k)e^{s\rho(j,k)}\,\rho(j,k). \tag{۲۳.۷}$$

چون $I(Q)$ می‌نیمم شده است

$$\begin{aligned} R(D)&=I(Q)\\ &=\sum_j\sum_k p(j)q(k|j)\log\frac{q(k|j)}{q(k)}. \end{aligned} \tag{۲۴.۷}$$

از

$$\frac{q(k|j)}{q(k)}=\lambda(j)e^{s\rho(j,k)} \tag{۲۵.۷}$$

نتیجه می‌شود

$$R(D)=s\,D\log e+\sum_j p(j)\log\lambda(j). \tag{۲۶.۷}$$

اکنون یک عبارت برای نرخ منبع به عنوان تابعی از دگرشکلی مجاز D به دست آوردیم. جواب به صورت ضمنی برحسب پارامتر s به دست آمده است، به استثنای چند حالت ساده رابطهٔ صریحی امکان‌پذیر نیست. یک مقدار s یک مقدار برای D و $R(D)$

می‌دهد و از این رو نقطه‌ای را روی منحنی $R(D)$ مشخّص می کند.

می‌توان نشان داد که این پارامتر s متناسب با مشتق اوّل $R(D)$ نسبت به D است و بنابراین به شیب منحنی $R(D)$ در نقطهٔ معیّنی مربوط می‌شود. داریم

$$s = \frac{dR(D)}{dD} / \log e. \qquad (۲۷.۷)$$

تابع نرخ دگرشکلی یک تابع یکنواخت کاهشی پیوسته برای $۰ \le D \le D_{\max}$ است. پارامتر s برای $۰ < D < D_{\max}$ پیوسته است و مثبت نمی‌باشد.

مثال ۴.۷

این مثال تعمیمی از مثالهای قبلی در این فصل است. مقدار تابع نرخ دگرشکلی را در حالت $D = ۰٫۴۵$ تعیین می کنیم. با جانشین کردن احتمالهای $q(k)$ (با مثال (۲.۷) مقایسه شود) و دگرشکلی (مثال (۱.۷) را ببینید) در معادلهٔ (۱۸.۷) نتیجه می‌شود

$$\lambda(۱) = \frac{۱}{\sum_k q(k) e^{s\rho(j,k)}} = \frac{۵}{۲ + e^s + ۲e^{۳s}}.$$

همین مقدار برای $\lambda(۲)$ مانند $\lambda(۱)$ به دست می‌آید. همچنین از $p(۱) = P(۲)$ و معادلهٔ (۲۳.۷) به دست می‌آوریم

$$\begin{aligned}
D &= \sum_{j=۱}^{۲} \sum_{k=۱}^{۳} \lambda(j) p(j) q(k) e^{s\rho(j,k)} \rho(j,k) \\
&= \lambda(۱) p(۱) \left\{ \sum_{k=۱}^{۳} q(k) e^{s\rho(۱,k)} \rho(۱,k) + \sum_{k=۱}^{۳} q(k) e^{s\rho(۲,k)} \rho(۲,k) \right\} \\
&= \frac{s/۲}{۲ + e^s + ۲e^{۳s}} \{۲(\frac{۱}{۵} e^s + \frac{۶}{۵} e^{۳s})\} \\
&= \frac{e^s + ۳e^{۳s}}{۲ + e^s + ۲e^{۳s}}.
\end{aligned}$$

تابع نرخ دگرشکلی به عنوان تابعی از D و s با کمک معادلهٔ (۲۶.۷) عبارت است از

$$\begin{aligned}
R(D) &= sD\log e + \sum_j p(j) \log \lambda(j) \\
&= sD\log e + \log \lambda(۱) \\
&= sD\log e + \log\{\frac{۵}{۲ + e^s + ۲e^{۳s}}\}.
\end{aligned}$$

با وجود این، علاقه‌مند به مقدار $R(D)$ برای $D = ۰٫۴۵$ هستیم. شکل (۴.۷) را ببینید.
با به کارگیری $x = e^s$ در عبارت مربوط به D نتیجه می‌شود

$$D = \frac{x + ۳x^۳}{۲ + x + ۲x^۳},$$

و از این رو

$$x + ۳x^۳ = ۰٫۴۵(۲ + x + ۲x^۳),$$

یا

$$۲٫۱x^۳ + ۰٫۵۵x - ۰٫۹ = ۰.$$

جواب این معادله عبارت است از $x \approx ۰٫۶۴$. بنابراین $s = \ln ۰٫۶۴ \approx -۰٫۴۵$.
اکنون مقدار $R(۰٫۴۵)$ را می‌توان به دست آورد.

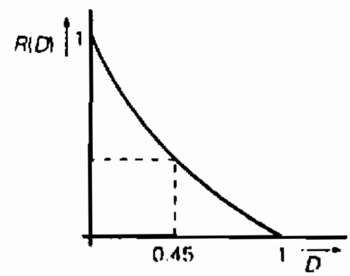
$$R(D) = s D \log e + \log\{\frac{۵}{۲ + e^s + ۲e^{۳s}}\}$$

$$= -۰٫۴۵ \times ۰٫۴۵ \times \log e + \log\{\frac{۵}{۲ + ۰٫۶۴ + ۲(۰٫۶۴)^۳}\}$$

$$= \frac{۱}{\ln ۲}(-۰٫۲۰ + ۰٫۴۶)$$

$$= \frac{۰٫۲۶}{\ln ۲} \approx ۰٫۳۷$$

$$\Rightarrow \quad R(۰٫۴۵) \approx ۰٫۳۷.$$ ▲



شکل ۴.۷- تابع نرخ دگرشکلی مثال (۴.۷)

تعیین تابع نرخ دگرشکلی معمولاً ساده نیست. به این علّت اغلب کران پایینی برای تابع نرخ دگرشکلی به کار برده می‌شود. روش عددی برای تعیین تابع $R(D)$ نیز وجود دارد.

۳.۷ حالت دودویی

یک کاربرد مهم مربوط به تابع نرخ دگرشکلی برای یک منبع دودویی با اندازهٔ دگرشکلی مذکور قبلی می‌باشد. در این حالت یک عبارت صریح ممکن است پیدا کرد، که به صورت زیر به دست می‌آید. فرض می‌کنیم که منبع، نمادها را با احتمالهای $p(u_1) = p$ و $p(u_2) = 1 - p$ تولید می‌کند. دگرشکلی همانند آنچه که در نمودار شکل (۱.۷) داده شده است می‌باشد.

فرض اضافی بدون کاستی در کلیت عبارت است از $0 \le p \le \frac{1}{2}$. به سادگی دیده می‌شود که $D_{\min} = 0$. این موضوع رخ می‌دهد اگر برای ماتریس انتقال Q را به صورت،

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

انتخاب کنیم. در این حالت تابع نرخ دگرشکلی $R(0)$ برابر با $H(U)$ است.

$$\begin{aligned} R(0) &= \min_{Q \in Q_D} I(Q) \\ &= \min_{Q \in Q_D} \{H(U) - H(U \mid \hat{U})\} \\ &= H(U) = -p \log p - (1-p) \log(1-p), \end{aligned}$$

چون ابهام صفر است.

ماکسیمم دگرشکلی $D = p$ است. این بهترین انتخابی است که می‌توان انجام داد اگر $I(Q) = 0$، زیرا حالت دیگر $D = 1 - p$ همواره دگرشکلی بزرگتری را می‌دهد چون $p \le \frac{1}{2}$.

ماتریس Q احتمالهای انتقال در این حالت برابر است با

$$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

برای این‌که نقاط دیگر منحنی $R(D)$ را تعیین کنیم نخست باید احتمالهای $q(k)$ را به دست آوریم.

از

$$\sum_j \lambda(j)\, p(j)\, e^{s\rho(j,k)} = 1,$$

و با فرض $a = e^s$، نتیجه می‌شود که

$$\lambda(0) \cdot p + \lambda(1) \cdot (1-p) \cdot a = 1,$$

$$\lambda(0) . p . a + \lambda(1) . (1-p) = 1,$$

بنابراین

$$\lambda(0) = \frac{1}{p(1+a)},$$

$$\lambda(1) = \frac{1}{(1-p)(1+a)}.$$

سپس $q(k)$ را از $\lambda(j)$ تعیین می‌کنیم. این کار را می‌توان براساس رابطهٔ زیر انجام داد

$$\lambda(j) = \frac{1}{\sum_k q(k) e^{s\rho(j,k)}}.$$

به دست می‌آوریم

$$q(0) + a q(1) = \frac{1}{\lambda(0)} = p(1+a),$$

$$a q(0) + q(1) = \frac{1}{\lambda(1)} = (1-p)(1+a),$$

از این نتیجه می‌شود

$$q(0) = \frac{p - a(1-p)}{1-a},$$

$$q(1) = \frac{1 - p - ap}{1-a}.$$

اگر این نتیجه را جانشین کنیم، برای دگرشکلی مجاز به دست می‌آوریم

$$D = \frac{a}{1+a}.$$

برای $R(D)$ داریم

$$R(D) = s D \log e + \sum_j p(j) \log \lambda(j)$$

چون

$$a = \frac{D}{1-D},$$

و

$$s = \frac{\log a}{\log e},$$

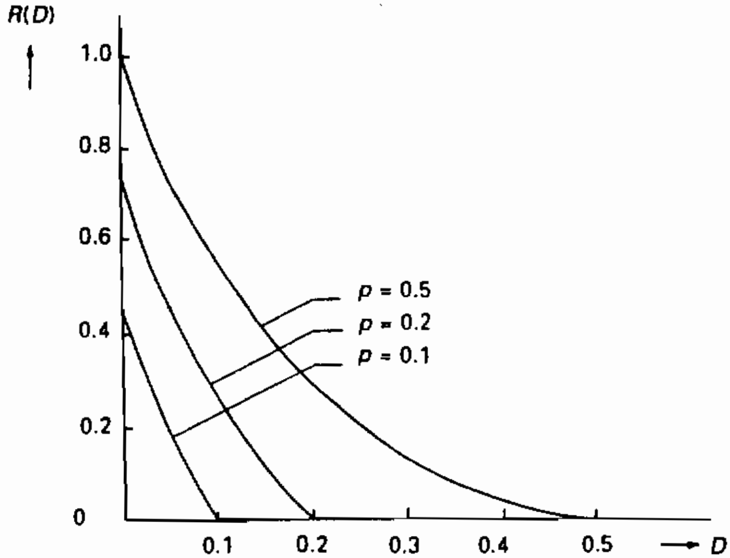
این عبارت را می‌توان به صورت زیر تبدیل کرد

$$R(D) = -p\log p - (۱-p)\log(۱-p) + D\log D + (۱-D)\log(۱-D)$$

$$= H(U) - H(D). \qquad (۲۸.۷)$$

بنابراین رابطه‌ای به دست آوردیم که از آن می‌توان مستقیماً اطلاع متقابل برای مقدار D معلوم که بایستی برای به دست آوردن متوسط دگرشکلی D ارسال گردد تعیین کرد.

شکل (۵.۷) منحنی $R(D)$ را برای چند مقدار p می‌دهد. از شکل می‌توان دریافت که تنها با افزایش $R(D)$ می‌توان متوسط دگرشکلی کوچکتری به دست آورد. همچنین، تابع نرخ دگرشکلی $R(D)$ با $p = ۰٫۵$ از تابع $p < ۰٫۵$ برای هر مقدار D بزرگتر است، که به طور شهودی درست است.



شکل ۵.۷- تابع نرخ دگرشکلی برای حالت دودویی

هر نقطه روی منحنی با ماتریس احتمالهای انتقال که به هر دو متوسط دگرشکلی $d(Q) = D$ و اطلاع متقابل $R(D)$ منجر می‌شود به دست می‌آید. برای احتمالهای انتقال داریم

$$q(k \mid j) = \lambda(j) q(k) e^{s\rho(j,k)},$$

که ماتریس Q زیر را می‌دهد:

$$Q = \begin{bmatrix} \dfrac{a(p-1)+p}{p(1-a^2)} & \dfrac{-a(p-1)-a^2p}{p(1-a^2)} \\ \dfrac{ap-a^2(1-p)}{(1-p)(1-a^2)} & \dfrac{(1-p)-ap}{(1-p)(1-a^2)} \end{bmatrix}. \qquad (۲۹.۷)$$

۴.۷ کدگذاری منبع و قضایای انتقال اطلاع

در این بخش پی خواهیم برد که قضیهٔ کدگذاری را برای حالتی که در آن متوسط دگرشکلی معیّنی مجاز است می‌توان به دست آورد. خواهیم دید که کدی با متوسط دگرشکلی D وجود دارد که اطلاع را ارسال می‌کند وقتی تابع نرخ دگرشکلی $R(D)$ کوچکتر یا حدّاقل برابر با C ظرفیت کانال است. بنابراین در یک روش گفتاری تابع نرخ دگرشکلی همان نقش مقدار اطلاع را بازی می‌کند. به عنوان یک نتیجه تابع نرخ دگرشکلی را می‌توان به عنوان مقدار اطلاع مؤثری که منبع اطلاع تولید می‌کند در نظر گرفت، اگر مقصد متوسط دگرشکلی D را مجاز بداند. در بخش قبل برای تولید مجدّد کامل به دست آوردیم که $R(D) = R(0) = H(U)$، بنابراین ممکن است مقدار اطلاع را به عنوان حالت خاصی از تابع نرخ دگرشکلی، یعنی برای $D = 0$، در نظر گرفت.

به ویژه مفهوم مقدار اطلاع مؤثّر مورد توجّه است زیرا اکنون قادریم آن را به این سؤال تبدیل کنیم که چه اطلاعی باید ارسال شود. در مورد سؤال بعدی یعنی چگونه باید این را تشخیص داد خیلی بیشتر شناخت داریم. سعی می‌کنیم مقدار اطلاع تولید شده توسط یک منبع را از طریق به کارگیری روشهای گوناگون تقلیل داده‌ها کاهش دهیم؛ مثلاً در مورد گفتار به قابلیّت درک در مقصد با سیگنالهای تصوّری، قابلیت تشخیص و غیره توجّه می‌شود. در پایان، این روشها به حذف اطلاعاتی که بی‌ارتباط با منبعند منجر می‌شود و اطلاع مؤثّر باقی می‌ماند. بنابراین در این روش یک مقدار دگرشکلی معیّنی از روی آگاهی معرّفی می‌شود به این طریق که مقصد بخش مناسبی از مقدار اطلاع تولید شده توسط منبع (تقلیل داده‌ها) مربوط به موضوع را دریافت می‌کند. برای رسیدن به فرمول‌بندی رابطهٔ ذکر شده بین $R(D)$ و C باید فرآیند تقلیل داده‌ها را با جزئیات بیشتری بررسی کنیم. چون منبع اطلاع اصلی را ممکن است به عنوان منبع جدید جایگزین شده در نظر گرفت، یعنی تولید مجدّد، که مقدار اطلاع مؤثّر معیّنی را تولید می‌کند، این فرآیند را ***کدگذاری منبع*** می‌نامیم.

واژهٔ منبع u و واژهٔ مقصد (مثلاً کدواژه) $\hat{u}$ را در نظر می‌گیریم که به ترتیب از

الفبای منبع و تولید مجدّد هر یک با L نماد انتخاب شده‌اند. یک گروه B از N کدواژهٔ مختلف را یک کد با اندازهٔ N و طول L می‌نامیم. برای هر کدواژهٔ u واژهٔ $\hat{u} \in B$ را به قسمی انتخاب می کنیم که برای آن دگرشکلی $\rho(u,\hat{u})$ بین u و $\hat{u}$ کوچکترین باشد. این واژهٔ مقصد را با $\hat{u}_B$ نمایش می‌دهیم. در این صورت دگرشکلی حاصل که برای یک واژهٔ منبع u داده شده عبارت است از

$$\rho(u,\hat{u}_B) = \min_{\hat{u}\in B} \rho(u,\hat{u}). \qquad (۳۰.۷)$$

این دگرشکلی نیز به گروه B بستگی دارد، گروه دیگری ممکن است مقدار دیگری را بدهد. اکنون متوسط دگرشکلی ρ_B که با وجود واژه منبع دلخواه u حاصل می‌شود عبارت است از

$$\rho_B = E[\rho(u,\hat{u}_B)] = \sum_u p(u) \min_{\hat{u}\in B} \rho(u,\hat{u}). \qquad (۳۱.۷)$$

نرخ کد R به صورت زیر تعریف می‌شود

$$R = \log N \text{ واژه / بیت },$$

یا

$$= \frac{1}{L} \log M \text{ نماد / بیت }. \qquad (۳۲.۷)$$

در واقع این ماکسیمم مقدار اطلاع بر نماد نسبت به یک کد منبع با اندازهٔ N و طول بلوکی L است، وقتی احتمالهای همهٔ کد واژه‌ها یکی باشند.

اکنون مفهوم ***D-مجاز بودن*** را معرّفی می کنیم. می گوییم که یک کد B، از این رو یک گروه کد واژه‌ها، D-مجاز است اگر $\rho_B \le D$. کوچکترین اندازهٔ کد D-مجاز نیز با اهمیت است. این را با $N(L,D)$ نمایش می‌دهیم، چون علاوه بر تعداد نمادها L و اندازه نیز به متوسط دگرشکلی Dمجاز بستگی دارد. اکنون به قضیه‌ای که به عنوان قضیهٔ ***کدگذاری منبع*** شناخته شده است، می‌رسیم.

قضیهٔ ۱.۷ (قضیهٔ کدگذاری منبع)

برای هر $\varepsilon > 0$ و هر $D \ge 0$ عدد صحیحی مانند L یافت می‌شود به طوری که یک کد $(D+\varepsilon)$-مجاز با طول بلوکی L و با نرخ کد $R < R(D)+\varepsilon$ وجود داشته باشد.

به عبارت دیگر، نابرابری

$$\frac{1}{L} \log N(L, D+\varepsilon) < R(D) + \varepsilon \qquad (۳۳.۷)$$

برای L به قدر کافی بزرگ برقرار است. ■

اثبات این قضیه را در این جا نخواهیم داد.

تنها وارون قضیهٔ کدگذاری منبع را ثابت خواهیم کرد که این قضیهٔ بیان می‌کند هیچ منبع D-مجازی با نرخ کمتر از $R(D)$ وجود ندارد.

قضیهٔ ۲.۷ (وارون قضیهٔ کدگذاری منبع)

هیچ کد D-مجاز نرخی کمتر از $R(D)$ ندارد. یعنی، برای همهٔ nها داریم

$$\frac{1}{L}\log N(L,D) \geq R(D). \tag{۳۴.۷}$$

برهان

اگر $B=(\hat{u}_1,\hat{u}_2,\ldots,\hat{u}_N)$ یک کد D-مجاز باشد، فرض می‌کنیم $I(\hat{U};U)$ متوسط اطلاع متقابل حاصل را نشان می‌دهد وقتی هر u به $\hat{u}\in B$ رمزی شده برای آن $\rho(u,\hat{u})$ می‌نیمم شده است، یعنی $\rho(u,\hat{u}_B)$ (با معادلهٔ (۳۰.۷) مقایسه کنید). چون رمزگذاری یک فرآیند قطعی است نتیجه می‌شود که $H(\hat{U}|U)=0$ و از این رو

$$I(\hat{U};U)=H(\hat{U})\leq \log N. \tag{۳۵.۷}$$

همچنین حالتی است که

$$I(\hat{U};U)=H(U)-H(U_1,\ldots,U_L|\hat{U}_1,\ldots,\hat{U}_L). \tag{۳۶.۷}$$

با کمک فرمولهای (۸.۳) و (۱۱.۳) این نتیجه را می‌دهد

$$\begin{aligned} H(U_1,\ldots,U_L|\hat{U}_1,\ldots,\hat{U}_L) &\leq \sum_{i=1}^{L} H(U_i|\hat{U}_1,\ldots,\hat{U}_L) \\ &\leq \sum_{i=1}^{L} H(U_i|\hat{U}_i). \end{aligned} \tag{۳۷.۷}$$

چون منبع بی‌حافظه است، در این حالت داریم

$$H(U_1,\ldots,U_L)=\sum_{i=1}^{L} H(U_i), \tag{۳۸.۷}$$

و از این رو با معادلهٔ (۳۵.۷) به دست می‌آوریم

$$\frac{1}{L}\left\{\sum_{i=1}^{L} H(U_i)-H(U_i|\hat{U}_i)\right\}\leq \frac{1}{L}\log N. \tag{۳۹.۷}$$

اگر D_i متوسط دگرشکلی باشد که با آن iامین نماد تولید مجدّد شده است در این

صورت

$$R(D_i) \le I(U_i;\hat{U}_i) = H(U_i) - H(U_i \big| \hat{U}_i). \qquad (۴۰.۷)$$

با ترکیب فرمولهای (۳۹.۷) و (۴۰.۷) نتیجه می‌شود

$$\frac{1}{L}\sum_{i=1}^{L} R(D_i) \le \frac{1}{L}\log N. \qquad (۴۱.۷)$$

چون B گروهی D-مجاز است $L^{-1}\sum_i D_i \le D$ برقرار می‌باشد. محدّب بودن $R(D)$ موجب می‌شود

$$R(D) \le R(\frac{1}{L}\sum_{i=1}^{L} D_i) \le \frac{1}{L}\sum_{i=1}^{L} R(D_i). \qquad (۴۲.۷)$$

با ترکیب فرمولهای (۴۱.۷) و (۴۲.۷) قضیه ثابت می‌شود. ■

از این رو قضیهٔ (۱.۷) بیان می‌کند یک کد منبع وجود دارد که رمزگذاری منبع اطلاع با متوسط دگرشکلی کمی بزرگتر از D را ممکن می‌سازد، که به موجب آن که منبع نرخی می‌دهد که به طور کاملاً دلخواهی به تابع نرخ دگرشکلی $R(D)$ نزدیک می‌شود. با وجود این، نرخ نمی‌تواند بدون دگرشکلی بیشتر کوچکتر از $R(D)$ بشود، چنان که از قضیهٔ (۲.۷) پیداست.

براساس این قضیه، می‌توان گفت که ممکن است $R(D)$ را به عنوان کوچکترین نرخ کد برای متوسط دگرشکلی D-مجاز در نظر گرفت. یعنی این نشانی از می‌نیمم تعداد کدواژه‌ها را می‌دهد.

قضیهٔ کدگذاری منبع تنها بیان می‌کند که یک کد امکان‌پذیر است ولی مستقیماً نشان نمی‌دهد که چگونه می‌توان آن را به دست آورد. اکنون از مثالی استفاده می‌کنیم که جنبه‌های گوناگون را شرح می‌دهد. منبع متقارن بی‌حافظهٔ دودویی با نمادهای ۰ و ۱ را در نظر بگیرید. برای اندازهٔ دگرشکلی، اندازهٔ دگرشکلی $\rho(u,\hat{u})$ را اختیار می‌کنیم که برای آن $\rho(u,\hat{u}) = 1$ اگر $u \ne \hat{u}$ و $\rho(u,\hat{u}) = 0$ اگر $u = \hat{u}$. اکنون یک روش کدگذاری منبع را در نظر می‌گیریم که در آن یک نماد دودویی (در یک مکان ثابت) در پیامی به طول L حذف می‌شود و مجدداً در طرف تولید مجدّد با حدس ساده‌ای با $p(0) = p(1) = \frac{1}{2}$ اضافه می‌شود.

پس از پذیرش سیگنال ارسال شده عدم‌حتمیت یا مقدار اطلاع درست ۱ بیت است. نماد حذف شده ۰ یا ۱ می‌باشد. این بدین معناست که بیت $H(U \big| \hat{U}) = 1$. در این صورت ماکسیمم مقدار اطلاع منتقل شده برابر است با

$$R^*(D) = H(U) - H(U \mid \hat{U})$$

$$= \log ۲^L - ۱ = L - ۱ \text{ بیت / پیام}$$

$$= \frac{L-۱}{L} \text{ بیت / نماد}. \qquad (۴۳.۷)$$

برای دگرشکلی داریم

$$D = \frac{۱}{L}(\frac{۱}{۲} \times ۰ + \frac{۱}{۲} \times ۱) = \frac{۱}{۲L},$$

و از این رو

$$R^*(D) = ۱ - ۲D. \qquad (۴۴.۷)$$

تابع نرخ دگرشکلی برای یک منبع دودویی با نمادهای هم‌احتمال برابر است با (با معادلهٔ (۲۸.۷) مقایسه کنید)

$$R(D) = H(U) - H(D) = ۱ - H(D). \qquad (۴۵.۷)$$

نتایج در شکل (۶.۷) نشان داده شده‌اند. به سادگی دیده می‌شود که تولید مجدّد بدون دگرشکلی تنها برای $L \to \infty$ امکان‌پذیر است چون $R^*(D) \to ۱$ و $D \to ۰$.

در عمل معمولاً روشهای ماهرانه‌تری به کار برده می‌شود.

قضیهٔ (۱.۷) را دوباره مورد توجّه قرار می‌دهیم. قضیه بیان می‌کند که یک کد $(D+\varepsilon)$-مجاز وجود دارد که نگاشتی از واژه‌های منبع u به کد واژه‌های $\hat{u}$ با دگرشکلی کوچکتر یا برابر با $(D+\varepsilon)$ را تضمین می‌کند. هم‌زمان نرخ رمزگذار خروجی حداکثر برابر است با

$$R = \frac{۱}{L} \log N(L, D+\varepsilon).$$

براساس همین قضیه می‌دانیم که ممکن است R با دقّت دلخواهی به $R(D)$ نزدیک شود. چون قضیهٔ (۱.۷) تضمین می‌کند که می‌توان $R < R(D) + \varepsilon$ ساخت، قضیهٔ دوم کدگذاری کانال شانون، قضیهٔ (۵.۴)، ایجاب می‌کند که رمزگذار منبع خروجی را می‌توان در رمزگشا در طرف دیگر کانال با احتمال دلخواه به شرط آن که کانال دارای ظرفیت $C > R(D) + \varepsilon$ باشد کشف کرد. این مطلب به قضیهٔ زیر منجر می‌شود.

قضیهٔ ۳.۷ (قضیهٔ ارسال اطلاع)

برای هر $\varepsilon > ۰$ خروجی یک منبع اطلاع را می‌توان با دگرشکلی حداکثر D در

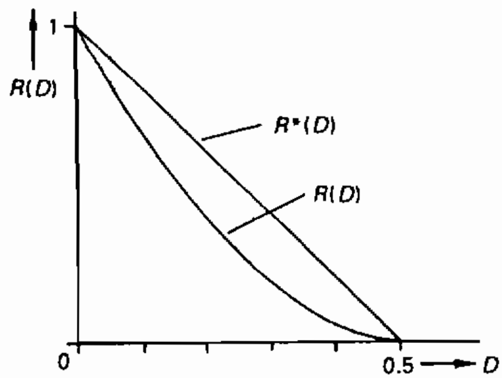
خروجی هر کانال بی‌حافظه گسسته با ظرفیت C دوباره تولید کرد، به شرط آن که

$$C > R(D) + \varepsilon. \qquad (۴۶.۷)$$ ■

از این رو می‌توان به جرأت گفت که سیستم ارتباطی به مفهوم نظریهٔ اطلاع ایده‌آل است اگر به متوسط دگرشکلی Dای برسد که برای این مقدار $R(D)$ داشته باشیم $R(D) = C$. این نتیجه دارای یک اثر مهم است، یعنی اکنون با یک منبع معلوم و کانالی معلوم با برابر قرار دادن تابع نرخ دگرشکلی منبع با ظرفیت کانال می‌توان تعیین کرد که متوسط دگرشکلی که به آن می‌توان رسید چیست. چون می‌خواهیم $R(D) \le C$ و چون $R(D)$ با افزایش D کاهش می‌یابد، این کوچکترین متوسط دگرشکلی ممکن نیز می‌باشد.

بنابراین یک کران پایینی برای متوسط دگرشکلی به دست آورده‌ایم که به نام ***کران نرخ دگرشکلی*** شناخته شده است.

در این جا رابطهٔ مستقیمی بین ویژگیهای رمزگذار، ظرفیت کانال و دگرشکلی معرّفی شده در طرف دریافت‌کننده ساخته می‌شود. همچنین قضیهٔ وارون ارسال اطلاع وجود دارد که بیان می‌کند تولید مجدّد منبع اطلاع با ماکسیمال دگرشکلی D-مجاز در طرف دریافت‌کننده هر کانالی با ظرفیت $C < R(D)$ غیرممکن است.



شکل ۶.۷- مقایسهٔ یک روش کدگذاری منبع با تابع نرخ دگرشکلی

از این رو اهمیت عملی در این است که اگر یک طراح سیستم بخواهد یک دگرشکلی حداکثر D داشته باشد و ظرفیت کانال C باشد تنها اگر $C \ge R(D)$ احتمالاً می‌تواند موفق شود.

۵.۷ تابع نرخ دگرشکلی پیوسته

مهمترین کاربرد نظریهٔ نرخ دگرشکلی در تولید مجدّد منبع اطّلاع پیوسته می‌باشد. دلیل آن این است که کاملاً امکان‌پذیر است که یک تولید مجدّد بدون دگرشکلی با یک منبع گسسته به دست آورد، در حالی که این کار بنابر تعریف برای منبع پیوسته غیرممکن و معمولاً نیز غیرضروری می‌باشد. مخصوصاً، با کمّی کردن سیگنالهای پیوسته موقعیتهایی پیش می‌آید که در آن یک مقدار کمّی معیّن دگرشکلی پذیرفته شده است برای این که تعداد بیتهای ضروری را پایین نگه دارد.

اکنون تابع نرخ دگرشکلی را برای حالت پیوسته با جزئیات بیشتری بررسی می‌کنیم. تشابه روشنی با حالت گسسته وجود دارد.

یک منبع بی‌حافظهٔ پیوسته با چگالی احتمال $p(x)$ را در نظر می‌گیریم. چگالی احتمال شرطی $\hat{x}$ به شرط x را $q(\hat{x}|x)$ و چگالی احتمال در طرف گیرنده را $q(\hat{x})$ می‌نامیم. $\rho(x,\hat{x})$ را برای اندازهٔ دگرشکلی انتخاب می‌کنیم که وقتی مقدار $\hat{x}$ را از مقدار x به دست می‌آوریم دگرشکلی را می‌دهد. انتخاب این اندازه بعداً شرح داده خواهد شد. اکنون ممکن است متوسط دگرشکلی را برای چگالی احتمال شرطی داده شده $q(\hat{x}|x)$ به صورت زیر تعریف کنیم

$$d(q)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x)q(\hat{x}|x)\rho(x,\hat{x})\,dx\,d\hat{x}. \tag{۴۷.۷}$$

برای متوسط اطلاع متقابل نیز داریم

$$I(q)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x)q(\hat{x}|x)\log\frac{q(\hat{x}|x)}{q(\hat{x})}\,dx\,d\hat{x}. \tag{۴۸.۷}$$

تابع نرخ دگرشکلی $R(D)$ می‌نیمم $I(q)$ است که برای آن دگرشکلی $d(q)$ حداکثر برابر دگرشکلی مجاز D است. برای سیستم پیوسته این به تعریف زیر منجر می‌شود

$$R(D)=\inf_{q\in q_D} I(q), \tag{۴۹.۷}$$

که در آن q_D مجموعهٔ توابع چگالی احتمال شرطی است که در نابرابری $d(q)\le D$ صدق می‌کند.

محاسبهٔ تابع نرخ دگرشکلی، به طور مشابه با حالت گسسته انجام می‌شود. در این صورت داریم

$$q(\hat{x}|x)=\lambda(x)q(\hat{x})e^{s\rho(x,\hat{x})}, \tag{۵۰.۷}$$

با

$$\lambda(x)=\frac{1}{\int_{-\infty}^{\infty} q(\hat{x})\,e^{s\rho(x,\hat{x})}\,d\hat{x}}, \tag{۵۱.۷}$$

که از آن نتیجه می‌شود

$$\int_{-\infty}^{\infty} \lambda(x)\,p(x)\,e^{s\rho(x,\hat{x})}\,dx=1. \tag{۵۲.۷}$$

سرانجام نتایج زیر را داریم

$$D=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(x)\,p(x)\,q(\hat{x})\,e^{s\rho(x,\hat{x})}\,\rho(x,\hat{x})\,dx\,d\hat{x}, \tag{۵۳.۷}$$

و

$$R(D)=sD\log e+\int_{-\infty}^{\infty} p(x)\log\lambda(x)\,dx. \tag{۵۴.۷}$$

بنابراین رابطهٔ بین $R(D)$ و D به دست آمده در این جا به صورت پارامتر است. پارامتر s با مشتق $R(D)$ و بنابراین با شیب منحنی $R(D)$ متناسب است.

تعدادی از ویژگیهای به دست آمده برای منبع گسسته معتبر باقی می‌مانند. ممکن است بیان کنیم که

$$D_{\min}=0, \tag{۵۵.۷}$$

$$D_{\max}=\inf_{\hat{x}}\,\hat{x}\int_{-\infty}^{\infty} p(x)\,\rho(x,\hat{x})\,dx. \tag{۵۶.۷}$$

به علاوه، $R(D)$ برای $0<D<D_{\max}$ یک تابع نامنفی به طور یکنواخت کاهشی است. یک تفاوت آن با منبع گسسته این است که منحنی $R(D)$ وقتی $D\to 0$ به بی‌نهایت می‌گراید. یک دگرشکلی کوچکتر به این معناست که مایلیم منبع را با جزئیات بیشتری دوباره تولید کنیم. در این صورت بیتهای بیشتری برای تشخیص یک مقدار لازم است و اینها باید منتقل شوند. این بدین معناست که وقتی $D\to 0$، $R(D)$ به بی‌نهایت می‌گراید. البته این یک عدم امکان فیزیکی است.

دگرشکلی که با منابع پیوسته معرّفی می‌شود تابعی از تفاضل بین مقادیر تولید شده در منبع و بازسازی شده در مقصد می‌باشد. اغلب یک اندازهٔ دگرشکلی به کار برده شده معیار مربع خطاست، که چنین نشان داده می‌شود

$$\rho(x,\hat{x}) = (x-\hat{x})^2. \qquad (۵۷.۷)$$

یک نتیجهٔ این تعریف این است که خطاهای بزرگ وزن زیادتری دارند. قدرمطلق خطا نیز به عنوان اندازهٔ دگرشکلی به کار برده می‌شود، یعنی

$$\rho(x,\hat{x}) = |x-\hat{x}|. \qquad (۵۸.۷)$$

بدون ارائهٔ نتیجه‌ای، یادآوری می‌کنیم که یک منبع اطلاع با چگالی احتمال گاوسی $N(0,\sigma^2)$، تابع نرخ دگرشکلی -اگر مربع خطا را به عنوان اندازهٔ دگرشکلی انتخاب کنیم- به صورت زیر داده می‌شود

$$\left.\begin{aligned} R(D) &= \frac{1}{2}\log\frac{\sigma^2}{D}\ \text{بیت / نماد} \quad , \quad 0 \le D \le \sigma^2 , \\ &= 0 \quad , \quad D > \sigma^2 . \end{aligned}\right\} \qquad (۵۹.۷)$$

که در آن σ^2 توان منبع اطلاع بر نماد می‌باشد. منحنی $R(D)$ در شکل (۷.۷) رسم شده است.

ماکسیمم دگرشکلی مجاز D_{max} را ممکن است به صورت زیر تعیین کرد. اگر هیچ اطلاعی ارسال نشده باشد بهترین انتخابی که دریافت کننده می‌تواند انجام دهد شامل انتخاب مقدار متوسط (در این جا برابر صفر) است زیرا این بزرگترین چگالی احتمال را دارد و بنابراین محتملترین مقدار ارسال شده است. خطایی که اکنون انجام می‌شود برابر است با $(x-\sigma)^2 = x^2$ و متوسط خطا $d(q)$ برابر است با $E\{(x-\sigma)^2\} = E\{x^2\} = \sigma^2$. در این صورت از این نتیجه می‌شود که D_{max}، یعنی می‌نیمم دگرشکلی برای $I(q) = R(D) = 0$، برابر با σ^2 است.

علاوه بر این دیده می‌شود که $R(D)$ اگر D کوچکتر و کوچکتر شود به بی‌نهایت افزایش پیدا می‌کند. البته این یک تجرید ریاضی است؛ در سیستمهای عملی D کرانی برابر با کمترین مقدار سیگنال اندازه‌پذیر خواهد داشت.

اگر منبع اطلاع مقیّد به پهنای باند W باشد، ممکن است منبع را با $2W$ نمونه مستقل در ثانیه تعیین کرد. از این رو داریم

$$\left.\begin{aligned} R(D) &= W\log\frac{\sigma^2}{D}\ \text{بیت / ثانیه} \quad , \quad 0 \le D \le \sigma^2 , \\ &= 0 \quad , \quad D > \sigma^2 . \end{aligned}\right\} \qquad (۶۰.۷)$$

گذشته مفهوم کران دگرشکلی را معرفی کردیم، متوسط دگرشکلی است که

برای آن $R(D)=C$. کوچکترین متوسط دگرشکلی است که امکان‌پذیر می‌باشد.

این کران نرخ دگرشکلی را برای حالتی به دست می‌آوریم که در آن یک منبع اطلاع گاوسی با پهنای باند W_x و توان σ_x^2 بر نمونه را به کانالی که دارای پهنای باند W_c و توسط نوفه گاوسی سفید جمعی با توان σ_n^2 بر نمونه مغشوش شده است متّصل می‌کنیم. برای منبع داریم

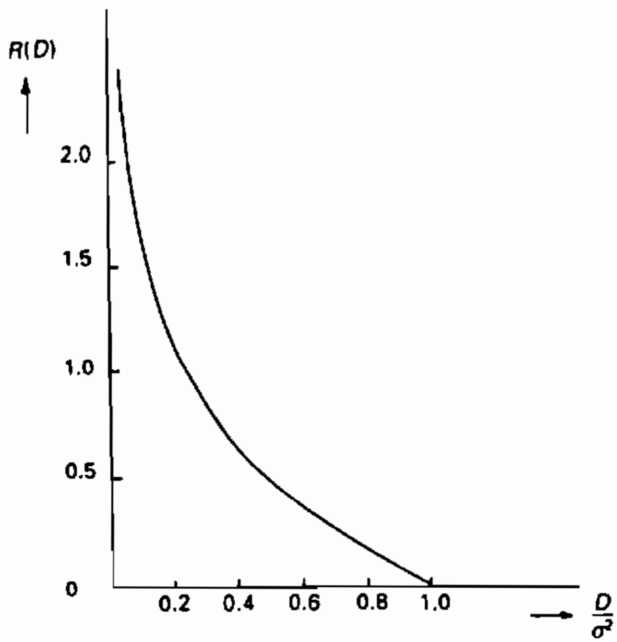
$$R(D)=W_x \log\frac{\sigma_x^2}{D} \text{ ثانیه / بیت} \qquad , \qquad 0 \le D \le \sigma_x^2 \tag{۶۱.۷}$$

ظرفیت کانال به صورت زیر است

$$C=W_c \log(1+\frac{P_x}{P_n}),$$

که در آن P_x توان سیگنال ورودی در کانال است. کران نرخ دگرشکلی با مساوی قرار دادن $R(D)$ و C به دست می‌آید. این نتیجه می‌دهد

$$D=\sigma_x^2\left\{1+\frac{P_x}{2W_c\sigma_x^2}\right\}^{-\frac{W_c}{W_x}} \qquad , \qquad 0 \le D \le \sigma_x^2. \tag{۶۲.۷}$$



شکل ۷.۷- تابع نرخ دگرشکلی برای حالت گاوسی

در نتیجه این کوچکترین دگرشکلی ممکن است که می‌توان به دست آورد. اگر حالت سازگار ایده‌آل $W_c = W_x = W$ و $P_x = ۲W\sigma_x^۲$ را در نظر بگیریم، در این صورت نتیجه می‌شود که

$$D' = \frac{\sigma_x^۲}{۱ + \frac{\sigma_x^۲}{\sigma_n^۲}}. \qquad (۶۳.۷)$$

اگر $\sigma_x^۲ << \sigma_n^۲$، آن‌گاه $D' \approx \sigma_x^۲$، که بدین معناست که کانال آن‌قدر بد است که می‌توان در عمل با انتخاب مکرّر میانگین مقدار سیگنال منبع انجام داد. اگر $\sigma_n^۲ << \sigma_x^۲$، آن‌گاه $D' \approx \sigma_n^۲$. این نیز روشن است که دریافت‌کننده نمی‌تواند سیگنال منبع اطلاع را با دقّت بیشتری از نوفهٔ مجاز در کانال پی ببرد زیرا $\sigma_n^۲$ کوچکترین توان محسوس در نمونه است. اگر منبع مستقیماً به کانال مرتبط شود، متوسط دگرشکلی برابر با این کوچکترین توان محسوس $\sigma_n^۲$ می‌باشد.

برای به دست آوردن تابع نرخ دگرشکلی همواره فرض می‌کنیم که $۰ \le D \le D_{max}$. در این مثال نشان داده شده است که برای $\sigma_x^۲ << \sigma_n^۲$ متوسط دگرشکلی به D_{max} نزدیک می‌شود به قسمی که بدون یک منبع می‌توان به عنوان امری واقعی همان نتیجه را در طرف گیرنده انتظار داشت. برای $\sigma_x^۲ >> \sigma_n^۲$، D به صفر میل نکرده بلکه به صورت $\sigma_n^۲$ باقی می‌ماند. برای این که قادر باشیم یک متوسط دگرشکلی $D < \sigma_n^۲$ به دست آوریم به نظر می‌رسد که می‌توان در جهت سیستمی تلاش کرد که در آن پهنای باند کانال بزرگ‌تر از پهنای باند منبع اطلاع است، از این رو $W_c > W_x$، که به این معناست که بایستی بسط - پهنای باند معینی را دنبال کرد. برای $W_c >> W_x$ متوسط دگرشکلی D به صفر میل خواهد کرد.

این مثال یک کاربرد کران نرخ دگرشکلی را شرح می‌دهد، که در آن می‌توان یک عبارت برای می‌نیمم متوسط دگرشکلی یافت که آن را با منبع معلوم و کانال معلوم به دست آورد. علاوه بر این، با به کارگیری عبارت یافته شده نشانی را پدید می‌آورد که می‌توان دریافت که چگونه می‌توان یک سیستم ارتباطی را اصلاح کرد.

۶.۷ تمرینها

۱.۷ یک منبع دودویی داده شده است که نمادهای $u_۰$ و $u_۱$ را از الفبای U تولید می‌کند. نمادهای منبع با احتمالهای $p(u_۰) = p$ و $p(u_۱) = ۱ - p$ $(۰ \le p \le \frac{۱}{۲})$ رخ می‌دهد.

کانال - Z همانند شکل (۸.۷) داده شده است.

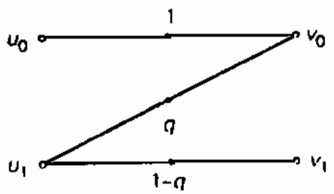
$\{v_0, v_1\}$ نمادهای مقصد از الفبای V هستند. احتمالهای انتقال $q(v_j|u_i)$ به صورت زیر داده شده است

$$q(v_0|u_0)=1 \quad , \quad q(v_1|u_1)=1-q.$$

دگرشکلی بین نمادهای منبع و تولید مجدد $\rho(u_i, v_j)$ به صورت زیر می‌باشند

$$\rho(u_0, v_0) = \rho(u_1, v_1) = 0,$$

$$\rho(u_1, v_0) = \rho(u_0, v_1) = 1.$$



شکل ۸.۷- کانال - Z تمرین (۱.۷)

(الف) متوسط دگرشکلی $d(Q)$ را محاسبه کنید.

(ب) تعریف تابع نرخ دگرشکلی را بدهید؛ ماکسیمم مقدار $R(D)$ چه قدر است؟ برای چه مقدار q این مقدار به دست می‌آید؟ متوسط دگرشکلی $d(Q)$ در این حالت چه قدر است؟

(پ) می‌نیمم مقدار $R(D)$ چه قدر است؟ برای چه مقدار q این مقدار به دست می‌آید؟ متوسط دگرشکلی $d(Q)$ در این حالت چه قدر است؟

(ت) $R(D)$ را به عنوان تابعی از D رسم کنید.

۲.۷ کانالی را با الفبای ورودی (x_1, x_2)، الفبای خروجی (y_1, y_2) و احتمالهای انتقال $q(y_j|x_i)$ در نظر بگیرید. در ورودی احتمال x_1 برابر $\frac{1}{4}$ است. برای دگرشکلی نماد داریم

$$\rho(x_1, y_1) = 0 \quad , \quad \rho(x_1, y_2) = \alpha \quad , \quad \rho(x_2, y_1) = 5-\alpha \quad , \quad \rho(x_2, y_2) = 0,$$

با $0 \le \alpha \le 5$.

در (الف) و (ب) فرض شده است که $q(y_1|x_1) = \frac{3}{5}$ و $q(y_1|x_2) = \frac{3}{10}$.

(الف) مقدار اطلاع در خروجی کانال را محاسبه کنید.

(ب) متوسط دگرشکلی را به عنوان تابعی از α محاسبه کنید. کوچکترین متوسط دگرشکلی قابل حصول چه قدر است؟

اکنون تابع نرخ دگرشکلی $r(D)$ سیستم داده شده را در نظر بگیرید.

(پ) $R(\circ)$ را محاسبه کنید و ماتریس کانال متناظر را به دست آورید.

(ت) $D_{\max}$ یعنی کوچکترین دگرشکلی ممکن را وقتی هیچ اطلاع از منبع دریافت نشده است محاسبه کنید. چه مقدار α بزرگترین $D_{\max}$ را خواهد داد؟

۷.۷ جوابها

۱.۷ (الف) برای احتمالهای توأم $p(u_i, v_j)$ به دست می‌آوریم

$$p(u_\circ, v_\circ) = p(v_\circ | u_\circ) p(u_\circ) = p,$$

$$p(u_\circ, v_۱) = p(v_۱ | u_\circ) p(u_\circ) = \circ,$$

$$p(u_۱, v_\circ) = p(v_\circ | u_۱) p(u_۱) = q(۱ - p),$$

$$p(u_۱, v_۱) = p(v_۱ | u_۱) p(u_۱) = (۱ - q)(۱ - p).$$

بنابراین متوسط دگرشکلی می‌شود

$$d(Q) = \sum_{i=\circ}^{۱} \sum_{j=\circ}^{۱} p(u_i, v_j) \rho(u_i, v_j)$$
$$= q(۱ - p).$$

(ب) برای تابع نرخ دگرشکلی در این حالت داریم

$$R(D) = \min_{Q \in Q_D} I(Q) \text{ نماد / بیت},$$

که در آن $I(Q)$ اطلاع متقابل است و

$$Q_D = \{Q | d(Q) \le D\}$$

تابع نرخ دگرشکلی به طور ماکسیمم برابر است با

$$\max R(D) = H(U)$$

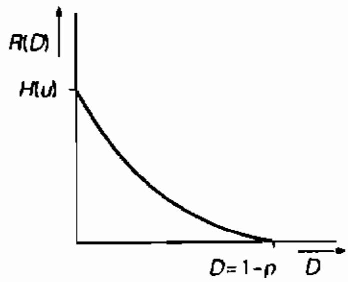
این ماکسیمم رخ می‌دهد، اگر یک رابطهٔ یک به یک بین نمادهای منبع و تولید مجدّد وجود داشته باشد. با توجّه به احتمالهای انتقال نتیجه می‌شود

$$q(v_0|u_0)=q(v_1|u_1)=1$$

و از این رو $1-q=1$ و $q=0$. اکنون، متوسط دگرشکلی برابر می‌شود با $d(Q)=0$.

(پ) مقدار مینیمال اطلاع متقابل $I(Q)$ برابر صفر است و از این رو مقدار مینیمال تابع نرخ دگرشکلی نیز برابر صفر است. این می‌نیمم اگر نمادهای تولید مجدّد هیچ اطلاعی دربارهٔ نمادهای منبع اصلی ندهند به دست می‌آید. این حالت را داریم اگر هر دو نماد منبع u_0 و u_1 به نماد تولید جدید یکسانی منجر شوند. این به $q=1$ منتهی می‌شود و برای متوسط دگرشکلی متناظر به دست می‌آوریم $d(Q)=1-p$.

(ت) شکل (۹.۷) را ببینید.



شکل ۹.۷- تابع نرخ دگرشکلی تمرین (۱.۷)

۲.۷ شکل (۱۰.۷) را ببینید.

توجّه: دگرشکلیهای دو به دو در داخل پرانتزها داده شده‌اند.

(الف) برای احتمالهای حاشیه‌ای $q(y_1)$ و $q(y_2)$ به دست می‌آوریم

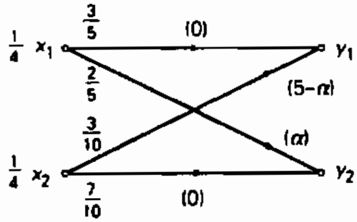
$$q(y_1)=\sum_{i=1}^{2}p(x_i)q(y_j|x_i)=\frac{1}{4}\times\frac{3}{5}+\frac{3}{4}\times\frac{3}{10}=\frac{3}{8},$$

$$q(y_2)=1-\frac{3}{8}=\frac{5}{8}.$$

اکنون مقدار اطلاع $H(Y)$ برابر می‌شود با

$$H(Y)=-\frac{3}{8}\log\frac{3}{8}-\frac{5}{8}\log\frac{5}{8}=0{,}954 \text{ بیت}.$$

(ب) متوسط دگرشکلی را به سادگی می‌توان به دست آورد



شکل ۷. ۱۰- دگرشکلیها و احتمالهای انتقال تمرین (۲.۷)

$$d(Q)=\sum_{i=۱}^{۲} p(x_i)p(y_j|x_i)\rho(x_i,y_j)$$

$$=\frac{۱}{۴}\times\frac{۲}{۵}\times\alpha+\frac{۳}{۴}\times\frac{۳}{۱۰}(۵-\alpha)=-\frac{۱}{۸}\alpha+\frac{۹}{۸}.$$

متوسط دگرشکلی برای $\alpha=۵$ دارای یک می‌نیمم است:

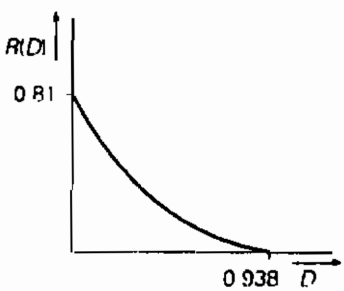
$$\min_{\alpha} d(Q)=\frac{۱}{۲}.$$

(پ) اگر $D=۰$، تابع نرخ دگرشکلی ماکسیممی برابر $H(X)$ را می‌دهد، که اطلاع منبع است،

$$H(X)=-\frac{۱}{۴}\log\frac{۱}{۴}-\frac{۳}{۴}\log\frac{۳}{۴}=۰{,}۸۱۱.$$

ماتریس متناظر کانال برابر است با

$$Q=\begin{bmatrix} ۱ & ۰ \\ ۰ & ۱ \end{bmatrix}$$



شکل ۱۱.۷- تابع نرخ دگرشکلی تمرین (۲.۷)

به این معنی که یک تناظر یک به یک بین نمادهای ورودی و خروجی وجود دارد.

(ت) براساس نظریه داریم

$$D_{\max} = \min_j \sum_{i=۱}^{۲} p(x_i)\rho(x_i, y_j).$$

در حالت موجود، به دست می‌آوریم

$$D_{\max} = \min(\frac{۱}{۴}\alpha, \frac{۳}{۴}(۵-\alpha))$$

و بنابراین

$$\begin{aligned} D_{\max} &= \frac{۱}{۴}\alpha \qquad , \qquad ۰ \le \alpha \le \frac{۱۵}{۴}, \\ &= \frac{۳}{۴}(۵-\alpha) \qquad , \qquad \frac{۱۵}{۴} \le \alpha \le ۵. \end{aligned}$$

به ازای $\alpha = \frac{۱۵}{۴}$، $D_{\max}$ ماکسیمم مطلق خودش ($= \frac{۱۵}{۱۶} \approx ۰٫۹۳۸$) را به دست می‌آورد.

۸
نظریهٔ اطلاع شبکه‌ای

۱.۸ مقدّمه

بحث ما تا کنون از ارتباط نقطه به نقطه تنها به حالتی مربوط می‌شود که در آن فرستنده و گیرنده توسط یک کانال ارتباطی خصوصی به هم متّصل می‌شوند. با وجود این، در عمل لزوم کاهش هزینه، فرستنده و گیرنده را مجبور خواهد کرد که از یک کانال ارتباطی با تعداد زیادی کاربر که می‌توانند به آن دسترسی داشته باشد استفاده نمایند.

به طور کلّی ***شبکه‌های ارتباطی چندپایانه‌ای*** شامل بیش از یک فرستنده یا گیرنده، یا حتّی بیش از یک کانال می‌باشد. مثالی از یک ***شبکه ارتباطی چند-مدخلی*** داده می‌شود که دارای چندین فرستنده و گیرنده است که همگی به یک کانال دسترسی دارند. اگر چندین فرستنده و تنها یک گیرنده وجود داشته باشد عبارت ***کانال ارتباطی چند-مدخلی*** به کار برده می‌شود. برای مثال، تعدادی ماهوارهٔ ایستگاه زمینی که همگی اطلاعاتشان را به یک ماهواره می‌فرستند. یکی از مهمترین جنبه‌های شبکهٔ ارتباطی چند-مدخلی آن است که مطمئن باشیم که کانال همواره به طور اپتیمال به کار برده می‌شود. از طرفی مناسب است که بهره‌برداری از ظرفیت کانال حداّکثر باشد، از طرف دیگر زمان دست‌یابی کاربر بایستی به قدر ممکن کوتاه باقی بماند. سر راست‌ترین پاسخ به این، کلید گردشی است. به استثنای یک فرستنده و گیرندهٔ تکی، کانال برای بقیهٔ کاربرها بسته است. از این رو، در واقع مسأله به یک ارتباط نقطه به نقطهٔ ساده تقلیل پیدا می‌کند. با وجود این، واضح است که این روش، فوق‌العاده کارا نیست. یک توقّف در حین ارتباط، بی‌درنگ موجب کاربرد

ناکارایی از کانال می‌شود. روش بهتری را با چند عضوی کردن می‌توان یافت کـــه در آن اطلاعات چندین منبع به یک گروه داده‌های تکی به هم پیوسته‌اند.

اگر تعدادی دریافت‌کننده و تنها یک فرستنده موجود باشد، در این صورت معمـــولاً جملهٔ ***کانال پخش*** به کار برده می‌شود. یک مثال مشـــهور فرستـــندهٔ تلویزیـــون اســـت کـــه برنامه‌هایی به گیرنده‌های تلویزیون در سراسر جهان می‌فرســـتد. یـــک نـــوع خـــاص کانـــال ارتباطی ***کانال دو-طرفه*** است. این کانال دو پایانه است که می‌تواند به طور هم‌زمان نقـــش فرستنده و گیرنده را بازی کند. این بدین معناست که گروه اطلاع در یک جهت منطبق بر گروه اطلاع در جهت دیگر خواهد بود.

مسألهٔ محاسبهٔ ظرفیت شبکه‌ها به طور سربسته هنوز یک مسألهٔ حل نشده است. ایـــن برای پیدا کردن کاراترین روش رمزگذاری نیز به کار می‌رود. حلّ رضایت‌بخشـــی بـــرای این مسائل را می‌توان تنها در حالتهای خیلی خاص معیّنی به دست آورد. نظریهٔ جامع دقیقی هنوز به دست نیامده است.

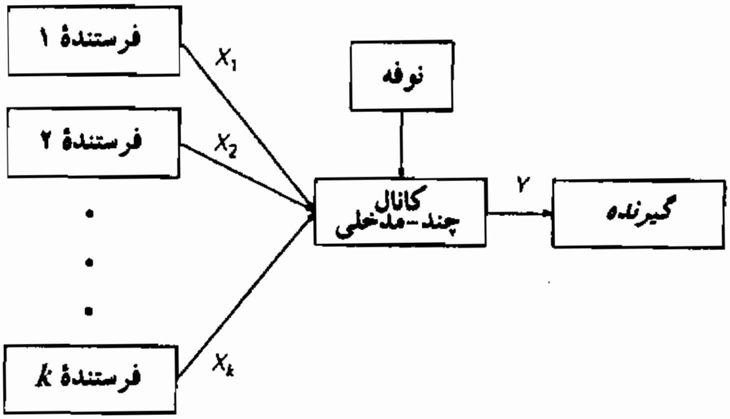
در این فصل جنبه‌های نظری اطلاع چندین شـــبکهٔ خـــاص را مـــانند کانـــال ارتبـــاطی چند-مدخلی کانال پخش و کانال دو-طرفه را بررسی خواهیم کرد.

۲.۸ کانال ارتباطی چند-مدخلی

یک کانال ارتباطی چند-مدخلی از تعدادی فرستنده و دقیقـــاً یـــک گـــیرنده تشـــکیل می‌شود، شکل (۱.۸) را ملاحظه کنید. همهٔ فرستنده‌ها یک کانال به کـــار می‌برنـــد. اطـــلاع دریافتی گیرنده برابر مجموع مثلاً k سیگنال فرستنده‌ها به اضافهٔ نوفهٔ معرّفـــی شـــده توســـط کانال می‌باشد. یک روش حلّ این مسأله، تقسیم کانال به تعدادی کانـــال فرعـــی و نســـبت دادن هر کانال فرعی به یکی از فرستنده‌ها می‌باشد. تقسیم کانـــال بـــه کانالهـــای فرعـــی را می‌توان با راههای مختلفی مثلاً با ***تقسیم فرکانس چند-مدخلی*** و ***تقسیم زمان چند-مدخلی*** انجام داد.

در حالت چند مدخلی به منظور تقسیم فرکانس، به هر کـــاربر یـــک بخـــش خـــوش تعریفی از باند فرکانس موجود برای فرستادن اطلاع به گیرنده اختصاص داده می‌شـــود. در معادلهٔ (۱۵.۶) عبارتی برای ظرفیت کانالی که متأثر از یک نوفهٔ گاوسی سفید جمعـــی بـــا توان متوسط P_n است به دست آوردیم. اگر P_x متوسط توان ورودی به کانال را نشان دهد، در این صورت

$$C = W \log\{1 + \frac{P_x}{P_n}\}. \qquad (۱.۸)$$



شکل ۱.۸- کانال ارتباطی چند-مدخلی

اگر W کلّ باند فرکانس به k بخش برابر تقسیم شود، در این صورت هر فرستنده می‌تواند یک پهنای باند $\frac{W}{k}$ را به جای W به کار برد. با قرار دادن مقدار $\frac{W}{k}$ در معادلهٔ (۱.۸) و یادآوری این که توان متوسط نوفه برای هر فرستنده با فاکتور k کاهش می‌یابد، عبارت زیر را برای ظرفیت کانال که هر کاربر می‌بیند به دست می‌آوریم:

$$C' = \frac{W}{k} \log\{1 + \frac{kP_x}{P_n}\}. \qquad (۲.۸)$$

برای کانال چند-مدخلی به منظور تقسیم زمان، هر کاربر ممکن است برای مدّت زمان از قبل تعیین شده به کانال دسترسی پیدا کند. بنابراین، با k کاربر، هر کاربر می‌تواند دقیقاً برای $\frac{1}{k}$ زمان از کانال استفاده نماید. در این مدّت زمان، هر کاربر کانال را با متوسط توان kP_x و بنابراین کانالی با ظرفیت C همانند آنچه که در عبارت (۲.۸) داده شده است می‌بیند.

همچنین ممکن است ظرفیت کانال را به هر یک از k کاربر بر طبق نرخ اطلاعات کاربرهای دیگر نسبت داد. اگر چندین کاربر نرخ کاهشی نشان دهند، در این صورت موقّتاً نرخ کاربرهای دیگر ممکن است افزایش یابد. طبیعتاً، نرخ R_i کاربر تکی i اگر بخواهد دسترسی خصوصی به کانال داشته باشد هرگز نمی‌تواند از ماکسیمم مقدار تجاوز کند. به عبارت دیگر

$$\mathrm{R}_i \leq W \log\left\{1+\frac{P_{x_i}}{P_n}\right\}. \tag{۳.۸}$$

اگر k کاربر باشند، در این صورت نرخ مجموع هرگز نمی‌تواند بزرگ‌تر از مجموع اطلاع گروهها با توان کلّ

$$\sum_{i=1}^{k} P_{x_i}.$$

باشد. بنابراین نتیجه می‌شود که

$$\sum_{i=1}^{k}\mathrm{R}_i \leq W \log\left\{1+\left(\sum_{i=1}^{k} P_{x_i} \Big/ P_n\right)\right\}. \tag{۴.۸}$$

اگر کاربران همکاری نکنند، در این صورت از نقطه نظر فرستندهٔ تکی فرستنده‌های دیگر را می‌توان به عنوان منابع نوفه در نظر گرفت. اگر این نوفه را بتوان به عنوان نوفهٔ گاوسی در نظر گرفت، کلّ توان نوفه برابر خواهد شد با

$$P_n + \sum_{j=1, j\neq i}^{k} P_{x_j},$$

ظرفیت کانال برابر است با:

$$C_i = W \log\left\{1+\left[P_{x_i} \Big/ \left(P_n + \sum_{j=1, j\neq i}^{k} P_{x_j}\right)\right]\right\}. \tag{۵.۸}$$

این عبارت برابر نرخ قابل دسترسی برای این حالت است.

اگر هر فرستنده در سطح توان یکسانی ارسال کند، یعنی برای هر j، $P_{x_j} = P_x$، به دست می‌آوریم

$$C = W \log\left\{1+\frac{P_x}{P_n + (k-1)P_x}\right\}. \tag{۶.۸}$$

چون داریم

$\ln a \leq a-1$ و بنابراین $\ln(a+1) \leq a$ و $\log(a+1) \leq a \log e$

نرخ هر کاربر محدود می‌شود به

$$\mathrm{R} \leq C = W \log\left\{1+\frac{P_x}{P_n + (k-1)P_x}\right\} \leq W \frac{P_x}{P_n + (k-1)P_x} \log e. \tag{۷.۸}$$

چنان که می‌توان انتظار داشت، وقتی سطح نوفه افزایش یابد یا چنانچه تعداد کلّ

کاربرها افزایش یابد، نرخ هر کاربر کاهش می‌یابد.

اگر تعداد کلّ کاربرها بزرگ باشد (k بزرگ است)، در این صورت

$$R \le \frac{W}{k} \log e. \qquad (۸.۸)$$

بنابراین نرخ هر یک از k کاربر باید در این عبارت صدق کند.

در بحث قبل حالت گاوسی را در نظر گرفتیم. اکنون با یک حالت کلّی‌تر که در آن فرض منبع گاوسی منظور نشده است اقدام می‌کنیم و می‌کوشیم عبارتی برای ظرفیت کانال چند-مدخلی یا به طور دقیقتر، نواحی ظرفیت را بیابیم.

با قضیه‌ای که شکل تعمیم‌یافتهٔ قضیهٔ پردازش داده‌ها (قضیهٔ (۶.۴)) است شروع می‌کنیم. این قضیه، نقش مهمی را در به دست آوردن شرایط لازم برای انتقال اطلاع از طریق یک کانال چند-مدخلی با احتمال خطای به قدر کافی کوچک بازی می‌کند.

قضیهٔ ۱.۸

مراحل پردازش متوالی را که در آن ورودی هر مرحلهٔ پردازش برابر خروجی مرحلهٔ پردازش قبلی است در نظر بگیرید (با شکل (۲.۸) مقایسه کنید). در این صورت اطلاع متقابل بین ورودی و خروجی با افزایش تعداد مراحل پردازش کاهش خواهد یافت:

$$I(U;V) \le I(X;Y). \qquad (۹.۸)$$

برهان

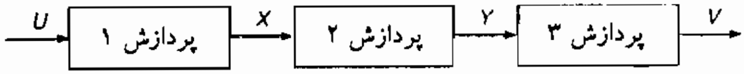
فرض کنید U، X، Y و Z ورودیها یا خروجیهای مراحل پردازش مختلف را نشان می‌دهند. بدون کاستی در کلیّت، ورودیها یا خروجیها را در سطح هر نماد در نظر خواهیم گرفت.

فرض کنید u_l، x_i، y_j و z_k نمایش نمادهای الفباهای متناظر باشند. جملهٔ (X,U) پیشامد توأم را نشان می‌دهد، که با احتمالهای توأم $p(x_i,u_l)$ مشخّص شده است. بنابراین، اطلاع متقابل بین Y و (X,U) برابر خواهد شد با

$$I(Y;(X,U)) = \sum_{i,j,l} p(u_l,x_i,y_j) \log\left\{\frac{p(y_j|x_i,u_l)}{p(y_j)}\right\}$$

$$= \sum_{i,j,l} p(u_l,x_i,y_j) \log\left\{\frac{p(y_j|x_i,u_l)}{p(y_j|x_i)}\right\} + \sum_{i,j,l} p(u_l,x_i,y_j) \log\left\{\frac{p(y_j|x_i)}{p(y_j)}\right\}. \qquad (۱۰.۸)$$

اولّین جملهٔ خط پایین معادلهٔ (۸ . ۱۰) را می‌توان به عنوان اطلاع متقابل بین U و Y،



شکل ۲.۸- مراحل پردازش متوالی

به شرط X، تفسیر کرد. در واقع جملهٔ دوم اطلاع متقابل بین X و Y می‌باشد.

$$I(Y;(X,U)) = I(U;Y|X) + I(X;Y).$$

با تعویض X و U، همچنین نتیجه می‌شود که

$$I(Y;(X,U)) = I(U;Y|X) + I(X;Y)$$
$$= I(X;Y|U) + I(U;Y). \qquad (۱۱.۸)$$

چون خروجی Y تنها به X وابسته است و به U بستگی ندارد، در این صورت داریم

$$\forall \quad i, j, l \qquad , \qquad p(y_j | x_i, u_l) = p(y_j | x_i).$$

با جایگزین کردن این عبارت در $I(U;Y|X)$ نتیجه می‌شود

$$I(U;Y|X) = 0. \qquad (۱۲.۸)$$

بنابراین:

$$I(Y;(X,U)) = I(X;Y) = I(X;Y|U) + I(U;Y) \qquad (۱۳.۸)$$

و

$$I(X;Y) \geq I(U;Y). \qquad (۱۴.۸)$$

به همین طریق می‌توان نشان داد که

$$I(U;Y) \geq I(U;V). \qquad (۱۵.۸)$$

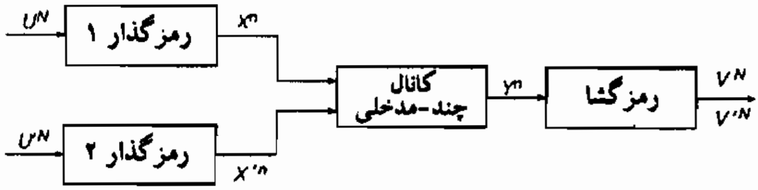
سرانجام، با ترکیب کردن این دو نابرابری نتیجه می‌شود

$$I(X;Y) \geq I(U;Y) \geq I(U;V), \qquad (۱۶.۸)$$

که با آن درستی قضیه نشان داده می‌شود . ■

قضیه بیان می کند که پردازش اطلاع موجب می‌شود اطلاع متقابل بین ورودی و خروجی کاهش یابد. به عبارت دیگر، پردازش موجب خواهد شد که ورودی و خروجی مستقلتر شوند.

اکنون وضعیت شکل (۳.۸) را در نظر بگیرید.



شکل ۳.۸- کانال چند-مدخلی با دو فرستنده

ورودی رمزگذار ۱ یک سری از بلوکهای U^N به طول N، بـا نمادهـای منتخب از الفبایی با اندازه J می‌باشد. خروجی رمزگشا یک سری از بلوکهای X^n به طول n است. به همین روش برای رمزگذار ۲ ورودی و خروجی بلوکهایی هستند که بـه ترتیب بـا U'^N و X'^n نمایش داده می‌شود. خروجی کانال که بی‌حافظه در نظر گرفتـه شـده اسـت یـک سری از کدواژه‌های Y^n می‌باشد. سرانجام، نتایج کدگشایی بلوکهـای خروجـی را بـا V^N و V'^N نمایش می‌دهیم که به ترتیب وابسته به بلوکهای اصلی U^N و U'^N می‌باشند.

با در نظر گرفتن یک کانال چند-مدخلی با دو فرستنده ***ناحیهٔ ظرفیت*** بـه صـورت غلاف محدّبی از همهٔ زوج نرخهای (R_1,R_2) است. که R_1 و R_2 به ترتیـب نرخهـای تکـی کاربر ۱ و ۲ را نشان می‌دهند، که برای آن کدهایی وجـود دارد کـه انتقـال اطـلاع را از طریق کانال با احتمال خطای کوچک دلخواهی ممکن می‌سازد. بدیهی است که اگر یکـی از نرخها برابر صفر باشد، آنگاه نرخ دیگر تنها با ماکسیمم ظرفیت کانال محدود می‌گردد. بنابراین مختصّات $(C_1,0)$ و $(0,C_2)$ در داخل ناحیهٔ ظرفیت قرار می‌گیرند. علاوه بـر ایـن، اگر $(C_1,0)$ و $(0,C_2)$ در ناحیهٔ ظرفیت واقع شوند، در این صـورت نقطـهٔ $(\lambda C_1,(1-\lambda)C_2)$ برای تمام مقادیر $\lambda\in[0,1]$ نیز بایستی در داخل ناحیه واقع شود. این مطلب با تشخیص ایـن که اگر تقسیم زمان به کار برده شود، کاربر ۱ برای کسر λ از زمـان موجـود بـه کانـال دسترسی خواهد داشت و کاربر ۲ ممکن است بـرای کسـر باقی‌مـانده $(1-\lambda)$ بـه کانـال دسترسی پیدا کند، روشن می‌شود. یعنی، نقاط زیر خطی که نقـاط $(C_1,0)$ و $(0,C_2)$ را بـه هم وصل می‌کند نیز نرخهایی را نشان می‌دهند که برای آنها یک احتمال خطای کوچـک دلخواهی می‌توان به دست آورد.

قضیهٔ زیر ما را قادر می‌سازد که به طور روشن کرانهای ناحیهٔ ظرفیت را تعریف کنیم.

قضیهٔ ۲.۸

فرض می‌کنیم که ورودیهای X و X' کانال چند-مدخلی متقابلاً مسـتقلّند، بنـابراین در این حالت برای هر توزیع توأم داریم

$$R_1 > I(X;Y|X') \tag{۱۷.۸}$$

یا

$$R_2 > I(X';Y|X) \tag{۱۸.۸}$$

یا

$$R_1 + R_2 > I((X;X');Y), \tag{۱۹.۸}$$

در این صورت هیچ کد با نرخ (R_1,R_2) وجود ندارد که بتواند احتمال خطای کوچک دلخواهی را پس از این که گیرنده اطلاعات را رمزگشایی کرده است تضمین نماید.

برهان

اثبات براساس نابرابری فانو می‌باشد.

اطلاعات را در سطح نماد در نظر می‌گیریم، رابطهٔ بین اطلاع شرطی نسبت به lامین نماد خروجی v_l رمزگشا به شرط lامین نماد ورودی u_l رمزگذار ۱ از یک طرف و احتمال خطا از طرف دیگر به صورت زیر داده می‌شود:

$$P_{e_1}\log(J-1) + H(P_{e_1}) \geq H(v_l|u_l), \tag{۲۰.۸}$$

که در آن زیرنویس رمزگذار ۱ را نشان می‌دهد. این موضوع مستقیماً از نابرابری فانو (قضیهٔ (۴.۴) را ببینید) و از این حقیقت که با وارونه کردن نقش X و Y در قضیهٔ (۴.۴) به سادگی کران یکسانی برای $H(Y|X)$ به دست می‌آید، نتیجه می‌شود.

با جمع کردن هر دو طرف روی تمام نمادهای بلوک و با توجّه به این امر که اندازه‌های اطلاع و احتمال خطا توابعی مقعّر هستند نتیجه می‌شود که

$$P_{e_1}\log(J-1) + H(P_{e_1}) \geq \frac{1}{N}\sum_{l=1}^{N} H(v_l|u_l). \tag{۲۱.۸}$$

به همین روش، نسبت به رمزگذار ۲، به دست می‌آوریم

$$P_{e_2}\log(J'-1) + H(P_{e_2}) \geq \frac{1}{N'}\sum_{l=1}^{N'} H(v'_l|u'_l). \tag{۲۲.۸}$$

واضح است برای بلوکهای U^N و V^N به طول N در این حالت داریم

$$H(V^N|U^N) \leq \sum_{l=1}^{N} H(v_l|u_l). \tag{۲۳.۸}$$

گرچه از طرف دیگر با بهره‌وری از تعریف احتمال متقابل می‌توان نوشت

$$H(V^N|U^N) = H(V^N) - I(U^N;V^N).$$

با به‌کارگیری قضیهٔ (۱.۸)، اکنون به دست می‌آوریم

$$\begin{aligned} H(V^N|U^N) &= H(V^N) - I(U^N;V^N) \\ &\geq H(X^n) - I(X^n;Y^n) \\ &\geq H(X^n) - I(X^n;(Y^n|X'^n)). \end{aligned} \quad (۲۴.۸)$$

در نابرابری اخیر، عبارت $I(X^n;(Y^n|X'^n))$ اطلاع متقابل بین خروجی رمزگذار ۱ و خروجی رمزگشا را نشان می‌دهد، با فرض این که خروجی رمزگذار ۲ معلوم است. نابرابری براساس این حقیقت است که

$$I(X^n;Y^n) \leq I(X^n;(Y^n|X'^n)). \quad (۲۵.۸)$$

این را می‌توان به صورت زیر شرح داد

فرض کنید x^n، x'^n و y^n بلوکهای مختلفی را که می‌تواند رخ دهد نشان می‌دهند. در این صورت نتیجه خواهد شد

$$I(X^n;Y^n|X'^n) = \sum_{x^n,x'^n,y^n} p(x^n,x'^n,y^n)\log\left\{\frac{p(y^n|x'^n,x^n)}{p(y^n|x'^n)}\right\}$$

$$= \sum_{x^n,x'^n,y^n} p(x^n,x'^n,y^n)\log\left\{\frac{p(y^n|x^n)}{p(y^n)}\right\}$$

$$+ \sum_{x^n,x'^n,y^n} p(x^n,x'^n,y^n)\log\left\{\frac{p(y^n)p(y^n|x'^n,x^n)}{p(y^n|x^n)p(y^n|x'^n)}\right\}$$

$$= I(X^n;Y^n) + \sum_{x^n,x'^n,y^n} p(x^n,x'^n,y^n)\log\left\{\frac{\dfrac{p(y_n)p(x^n|x'^n,y^n)p(x'^n,y^n)}{p(x'^n,x^n)}}{\dfrac{p(y^n|x^n)p(x'^n,y^n)}{p(x'^n)}}\right\}$$

$$= I(X^n;Y^n) + \sum_{x^n,x'^n,y^n} p(x^n,x'^n,y^n)\log\left\{\frac{\dfrac{p(y^n)p(x^n\,|\,x'^n,y^n)}{p(x'^n,x^n)}}{\dfrac{p(y^n,x^n)}{p(x'^n)p(x^n)}}\right\}$$

$$= I(X^n;Y^n) + \sum_{x^n,x'^n,y^n} p(x^n,x'^n,y^n)\log\left\{\frac{p(x^n\,|\,x'^n,y^n)}{p(x^n\,|\,y^n)}\right\}$$

$$= I(X^n;Y^n) + I(X'^n;X^n\,|\,Y^n).$$

بنابراین

$$I(X^n;Y^n\,|\,X'^n) \ge I(X'^n;Y^n).$$

با ترکیب معادلات (۲۳.۸) و (۲۴.۸) و توجّه به این که نرخ رمزگذار ۱ با $R_1 = \frac{H(X^n)}{n}$ داده شده است، به دست می‌آوریم

$$H(V^N\,|\,U^N) \ge H(X^n) - I(X^n;Y^n\,|\,x'^n)$$

$$= \sum_{i=1}^{N} H(V_l,U_l) \ge nR_1 - I(X^n;Y^n\,|\,X'^n)$$

$$\ge nR_1 - \sum_{l=1}^{n} I(X_l;Y_l\,|\,X'_l)$$

$$= \sum_{l=1}^{n}[R_1 - I(X_l;Y_l\,|\,X'_l)]. \qquad (۲۶.۸)$$

با تکرار این مراحل برای رمزگذار ۲ به نابرابری زیر خواهیم رسید

$$\sum_{i=1}^{N} H(V'_l,U'_l) \ge \sum_{l=1}^{n}[R_2 - I(X'_l;Y_l\,|\,X_l)]. \qquad (۲۷.۸)$$

مجموع این دو عبارت به صورت زیر است

$$\sum_{l=1}^{N} H(V_l,U_l) + \sum_{i=1}^{N} H(V'_l,U'_l) \ge \sum_{l=1}^{n}[R_1 + R_2 - I((X_l,X'_l);Y_l)]. \qquad (۲۸.۸)$$

بنابراین، با کمک نابرابری فانو، اکنون سه نابرابری زیر را داریم

$$N\{P_{e_1}\log(J-1)+H(P_{e_1})\}\geq\sum_{l=1}^{n}[\mathsf{R}_1-I(X_l;Y_l|X'_l)], \quad (۲۹.۸)$$

$$N'\{P_{e_2}\log(J'-1)+H(P_{e_2})\}\geq\sum_{l=1}^{n}[\mathsf{R}_2-I(X'_l;Y_l|X_l)], \quad (۳۰.۸)$$

$$N\{P_{e_1}\log(J-1)+H(P_{e_1})\}+N'\{P_{e_2}\log(J'-1)+H(P_{e_2})\}$$
$$\geq\sum_{l=1}^{n}[\mathsf{R}_1+\mathsf{R}_2-I((X_l,X'_l);Y_l)]. \quad (۳۱.۸)$$

اگر یکی از عبارتهای سمت راست علامت بزرگتر یا برابر برای هر مقدار I مثبت باقی بماند، در این صورت احتمال خطا نمی‌تواند صفر شود. به عبارت دیگر، امکان ندارد اطلاع را بدون خطا کدگشایی کرد. این بیان با آنچه که قضیه ایجاب می‌کند یکی است. ■

قضیهٔ وارون را می‌توان بدون مشکل خیلی زیادی فهمید و بنابراین اثبات کامل آن را حذف می‌کنیم.

قضیهٔ ۳.۸

فرض می‌کنیم n طول بلوک به قدر کافی بزرگ باشد، در این صورت کد مناسبی برای کانال چند-مدخلی با دو فرستنده با نرخهای R_1 و R_2 و احتمال خطای به قدر کافی کوچک می‌توان یافت، به شرطی که برای توزیع توأمی از ورودیهای X و X'، با X و X' متقابلاً مستقل، داشته باشیم

$$\mathsf{R}_1\leq I(X;Y|X'), \quad (۳۲.۸)$$

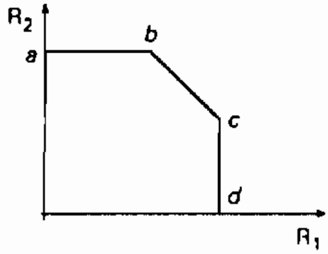
$$\mathsf{R}_2\leq I(X';Y|X), \quad (۳۳.۸)$$

و

$$\mathsf{R}_1+\mathsf{R}_2\leq I((X,X');Y). \quad (۳۴.۸)$$ ■

با در نظر گرفتن قضایای (۲.۸) و (۳.۸) می‌توان نتیجه گرفت که ناحیهٔ ظرفیت کانال بی‌حافظه سطح محصور شده توسط مرز محدّب تمام نرخهای R_1 و R_2 را که در سه شرط قضیهٔ (۳.۸) صدق می‌کنند برای توزیع احتمال حاصل‌ضرب معیّنی از زوج ورودی (X,X') می‌پوشاند.

برای توزیع احتمال حاصل‌ضرب خاصّی، ناحیهٔ ظرفیت شبیه شکل (۴.۸) خواهد بود.



شکل ۴.۸- ناحیهٔ ظرفیت قابل دسترسی یک کانال چند-مدخلی

نقطهٔ a متناظر با وضعیتی است که فرستندهٔ ۱ هیچ اطلاعی را به گیرنده نمی‌فرستد و فرستندهٔ ۲ اطلاع را با ماکسیمم نرخ می‌فرستد. این ایجاب می کند که:

$$\max R_2 = \max_{X,X'} I(X';Y|X).$$

برای هر توزیع حاصل‌ضرب در محدودهٔ (X,X') داریم

$$I(X';Y|X) = \sum_x p(x) I(X';Y|X=x)$$
$$\le \max_x I(X';Y|X=x).$$

این ماکسیمم را با جست‌وجوی توزیع X'ای که برای آن اطلاع متقابل ماکسیمم خودش را اختیار می کند می‌توان به دست آورد.

بنابراین:

$$\max R_2 \le \max_{X'} \max_x I(X';Y|X=x). \qquad (۳۵.۸)$$

نقطهٔ b متناظر با نرخ ماکسیمم است که در آن فرستندهٔ ۱ می‌تواند اطلاع خودش را به گیرنده بفرستد، اگر فرستندهٔ ۲ قبلاً با نرخ ماکسیمم ممکن در حال ارسال باشد. فرستندهٔ ۱ را می‌توان به عنوان اطلاعی که نوفه توسط فرستندهٔ ۲ به گیرنده ارسال می کند در نظر گرفت. بحث قبلی ما نشان می‌دهد که در این حالت فرستندهٔ ۱ نرخ ماکسیممی که می‌تواند کسب کند برابر $I(X;Y)$ است.

همین مطالب برای نقاط c و d به کار می‌رود، وقتی نقش فرستندهٔ ۱ و فرستندهٔ ۲ تعویض شوند.

همان‌طور که قبلاً اشاره شد توصیف حالت گاوسی از قضایایی که برای حالت کلّی داده شده‌اند با تفسیر کردن این قضایا برای مقادیر پیوسته بی‌درنگ نتیجه می‌شود. سیگنالی که به گیرنده‌ها می‌رسد برابر است با $\mathbf{y} = \mathbf{x} + \mathbf{x}' + \mathbf{n}$، که در آن $\mathbf{n}$ نوفهٔ گاوسی با توان P_n را

نشان می‌دهد. در نتیجه جملهٔ $I(X;(Y|X'))$ قضیهٔ (۳.۸) را می‌توان به صورت زیر نوشت:

$$\begin{aligned} I(X;Y|X') &= H(Y|X') - H(Y|X,X') \\ &= H(X+X'+\mathrm{N}|X') - H(X+X'+\mathrm{N}|X,X') \\ &= H(X+\mathrm{N}|X') - H(\mathrm{N}|X,X'). \end{aligned} \tag{۳۶.۸}$$

چون نوفه کلاً از X و X' مستقل است، بایستی نتیجه شود که:

$$H(\mathrm{N}|X,X') = H(\mathrm{N}). \tag{۳۷.۸}$$

به علاوه، استقلال متقابل X و X' ایجاب می‌کند که:

$$H(X+\mathrm{N}|X') = H(X+\mathrm{N}). \tag{۳۸.۸}$$

این نتیجه می‌دهد

$$\begin{aligned} I(X;Y|X') &= H(X+\mathrm{N}) - H(\mathrm{N}) \\ &= H(X+\mathrm{N}) - \log\sqrt{۲\pi eP_n} \\ &\le \log\sqrt{۲\pi e(P+P_n)} - \log\sqrt{۲\pi eP_n}, \end{aligned} \tag{۳۹.۸}$$

که در آن P توان قید بر فرستندهٔ ۱ است.

آخرین مطلب این امر را منعکس می‌کند که اندازهٔ اطلاع توزیع گاوسی همواره برابر ماکسیمم مقدار ممکن است.

با بازنویسی جملهٔ سمت راست به اختصار

$$I(X;Y|X') \le \frac{۱}{۲}\log\left(۱+\frac{P}{P_n}\right),$$

و نهایتاً به

$$\mathrm{R}_۱ \le \frac{۱}{۲}\log\left(۱+\frac{P}{P_n}\right). \tag{۴۰.۸}$$

منجر می‌شود.

با همین روش عباراتی برای قیود باقی‌ماندهٔ قضیهٔ (۳.۸) می‌توان به دست آورد:

$$\mathrm{R}_۲ \le \frac{۱}{۲}\log\left(۱+\frac{P'}{P_n}\right) \tag{۴۱.۸}$$

که P' برابر توان قید بر فرستندهٔ ۲ می‌باشد و

$$R_1 + R_2 \le \frac{1}{2}\log\left(1+\frac{P+P'}{P_n}\right). \qquad (۴۲.۸)$$

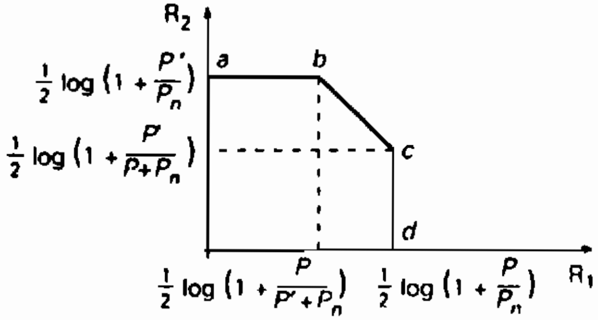
محاسبات بالا همگی برحسب بیت بر ارسال بیان می‌شوند. با وجود این، سیگنالهای پیوسته، که برای مدّت T طول می‌کشد با $2WT$ نمونه می‌توان نشان داد وقتی W پهنای باند سیگنال است، در نتیجه $2W$ نمونه در ثانیه اختیار می‌شود. با ضرب $2W$ در طرف راست فرمولهای (۸. ۴۰) تا (۴۲.۸) کرانهای بالایی برای نرخهای برحسب بیت بر ثانیه به دست خواهد آمد. این متناظر با فرمولهای (۳.۸) و (۴.۸) می‌باشد.

در عبارتهای (۴۲.۸) و (۴۴.۸) هنگامی که X و X' با توزیعهای گاوسی توصیف شده باشند برابری برقرار می‌باشد.

این سه عبارت تعریف ناحیهٔ ظرفیت شکل (۴.۸) را برای کانال چند-مدخلی گاوسی میسّر می‌سازند. اگر فرستندهٔ ۱ باید هیچ اطلاعی را اصلاً نفرستد، یعنی $R_1 = 0$، در این صورت نرخ ماکسیمم فرستندهٔ ۲ برابر $\frac{1}{2}\log(1+\frac{P'}{P_n})$ خواهد بود. به نقطهٔ a از شکل (۵.۸) مراجعه کنید. برعکس برای نقطهٔ d شکل. نرخهای مربوط به b و c را می‌توان به صورت زیر یافت. نخست، گیرندهٔ پیام ارسالی توسط فرستندهٔ ۳ را با توجّه به اطلاع ارسالی با فرستندهٔ ۱ به عنوان نوفه کدگشایی می‌کنند. نرخی که هنوز هم یک احتمال خطای کوچک دلخواه را مجاز می‌داند برابر است با

$$R_2 \le \frac{1}{2}\log\left\{1+\frac{P'}{P+P_n}\right\}. \qquad (۴۳.۸)$$

در این صورت پیام کدگشایی شدهٔ فرستندهٔ ۲ را از پیام دریافت‌شده کسر می‌کنیم، پس از آن پیام فرستندهٔ ۱ را می‌توان کدگشایی کرد.



شکل ۵.۸- ناحیهٔ ظرفیت برای کانال چند-مدخلی گاوسی با دو فرستنده

این، روش معقولی است، به شرطی که

$$R_1 \leq \frac{1}{2}\left\{1+\frac{P}{P_n}\right\}. \qquad (44.8)$$

این دقیقاً برابر با مختصّات گوشهٔ c است. مختصات نقطهٔ b را به طریق مشابه می‌توان به دست آورد.

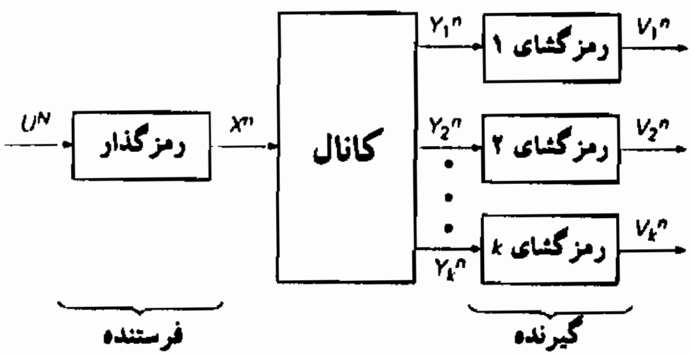
۳.۸ کانالهای پخش

یک کانال پخش شامل یک فرستنده و چندین گیرنده است. مثالی از این نوع کانال با فرستندهٔ تلویزیون یا رادیو داده می‌شود، که برنامه‌هایی پخش می‌کنند که توسط تعداد زیادی گیرنده دریافت می‌شوند. به علّت نوفه تفکیک‌نشدنی هیچ دو گیرنده‌ای واقعاً سیگنالهای یکسانی دریافت نمی‌کنند. شکل (۶.۸) را در نظر بگیرید که یک کانال پخش با k گیرنده را توصیف می‌کند.

در ادامهٔ بحث فرض می‌کنیم که تنها دو گیرنده وجود دارد: $k = 2$.

فرستنده، پیامهای X^n به طول n ارسال خواهد کرد که وابسته به پیامهای اصلی U^N به طول N می‌باشند. با در نظر گرفتن حضور نوفه در کانال کدگشای ۱، Y_1^n به جای X^n دریافت می‌کند.

پس از این که این پیام کدگشایی شد گیرنده V_1^N را دریافت می‌کند که فرم تغییر شکل یافته U^N می‌باشد. به همین منوال کدگشای ۲، Y_2^n دریافت خواهد کرد و این را به V_2^N کدگشایی می‌کند.



شکل ۶.۸- کانال پخش

در واقع، یک توصیف کلّی از کانال با احتمالهای انتقال $p(y_1, y_2|x)$ از الفبای رمزگذار خروجی به کانال ورودی داده می‌شود.

یک ***کانال پخش تضعیف شده*** کانال پخشی است که برای آن داریم

$$p(y_1, y_2|x) = p(y_1|x)p(y_2|x). \qquad (۴۵.۸)$$

کانال پخش گاوسی مثالی از کانال پخش تضعیف شده است. می‌توان نوشت $\mathbf{y}_1 = \mathbf{x} + \mathbf{n}_1$ که $\mathbf{n}_1$ نوفهٔ گاوسی با توان P_{n_1} را نشان می‌دهد و البته همچنین $\mathbf{y}_2 = \mathbf{x} + \mathbf{n}_2$. این عبارت را می‌توان به صورت زیر نوشت

$$\mathbf{y}_2 = \mathbf{x} + \mathbf{n}_2 = \mathbf{y}_1 + \mathbf{n}_3,$$

که در آن $\mathbf{n}_3$ متغیّر تصادفی گاوسی با توان $P_{n_2} - P_{n_1}$ را نشان می‌دهد. نمایش خلاصه‌ای از این عبارت در شکل (۷.۸) داده شده است.

مسلّماً کانال پخش را می‌توان به صورت تعدادی گیرنده که به طور سری به هم متّصل شده‌اند و پیامها را به یکدیگر منتقل می‌کنند در نظر گرفت. پیامهای ارسال شده در هر مرحله کمی ضعیف می‌شود.

ترکیب داده شده در شکل (۶.۸) داده شده با $k = ۲$ را در نظر بگیرید، در این صورت حالت گاوسی معادلات زیر را برای ظرفیتهای تکی خواهد داد:

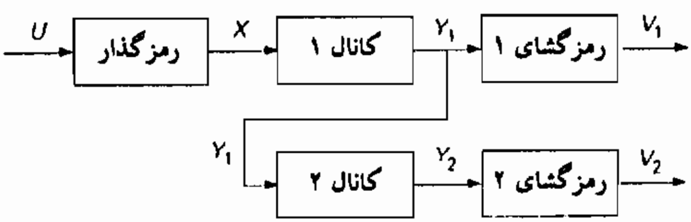
$$C_1 = W \log\{1 + \frac{P_x}{P_{n_1}}\} \qquad (۴۶.۸)$$

و

$$C_2 = W \log\{1 + \frac{P_x}{P_{n_2}}\}, \qquad (۴۷.۸)$$

که در آن P_x توان ورودی کانال می‌باشد.

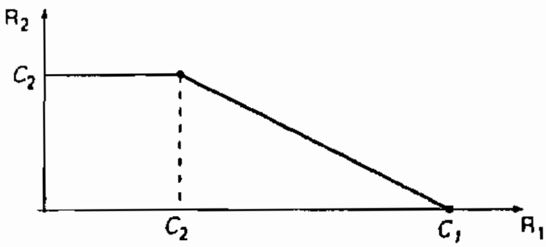
در بحث بعدی فرض می‌کنیم که $P_{N_1} < P_{N_2}$ و بنابراین $C_1 > C_2$.



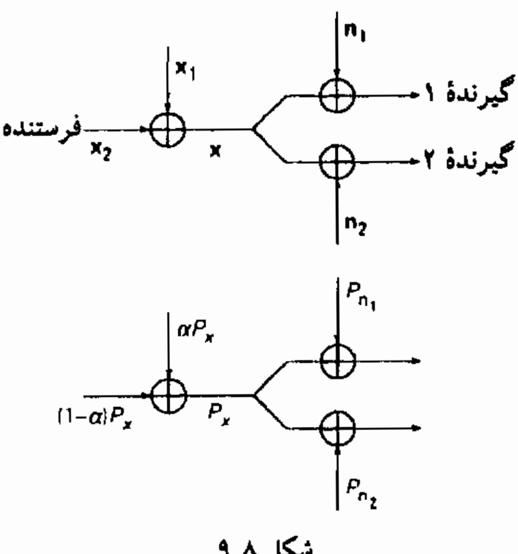
شکل ۷.۸- کانال پخش ضعیف شده

بدیهی است که تقسیم زمان جواب خاصّی را پیشنهاد می‌کند. در این حالت به ترتیب ابتدا کانال را برای ارسال پیامها به رمزگذار ۱ و سپس به رمزگذار ۲ آماده می‌کنیم. تقسیم زمان متناظر با نرخها در شکل (۸.۸) رسم شده‌اند.

با وجود این، روشهایی برای توسعهٔ اجرای کانال وجود دارند. فرض کنید پیام دادهٔ شدهٔ $\mathbf{x}$ با دو پیام $\mathbf{x}_1$ و $\mathbf{x}_2$ به ترتیب با توانهای مربوطه αP_x و $(1-\alpha)P_x$ که در آن $0 \le \alpha \le 1$ تطبیق می‌کنند، که در آن $\mathbf{x}_2$ می‌باید توسط گیرندهٔ ۲ با نوفهٔ بالا دریافت شود. از این رو گیرنده‌های ۱ و ۲ به ترتیب $\mathbf{y}_1 = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{n}_1$ و $\mathbf{y}_2 = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{n}_2$ را دریافت می‌کنند. شکل (۹.۸) را ببینید.



شکل ۸.۸- ناحیهٔ ظرفیت برای حالت تقسیم زمان



شکل ۹.۸

گیرندهٔ ۲، P_{n_2} و αP_x را به عنوان منابع نوفه که توان کلّ آن برابر $\alpha P_x + P_{n_2}$ است در نظر می‌گیرد. در نتیجه، پیامها را می‌توان به گیرندهٔ ۲ با احتمال خطای به قدر کافی

کوچک فرستاد به شرط آن که نرخ ارسال از مقدار زیر کوچکتر باشد

$$C_۲(\alpha) = W \log\left\{ ۱ + \frac{(۱-\alpha) P_x}{\alpha P_x + P_{n_۲}} \right\}. \qquad (۴۸.۸)$$

چون $P_{n_۱} < P_{n_۲}$، گیرندهٔ ۱ نیز قادر خواهد بود پیام $\mathbf{x}_۲$ را با احتمال خطایی که به قدر کافی کوچک است دریافت کند.

پس از کدگشایی $\mathbf{x}_۲$، گیرندهٔ ۲ می‌تواند $\mathbf{x}_۲$ را از $Y_۱$ کسر کند، نتیجه می‌شود

$$\mathbf{y}_۱ - \mathbf{x}_۲ = \mathbf{x}_۱ + \mathbf{n}_۱ .$$

اکنون گیرندهٔ ۱ باید از طریق کانالی با نوفهٔ گاوسی جمعی و توان ورودی αP_x مرتبط شود. ظرفیت این کانال برابر است با

$$C_۱(\alpha) = W \log\left(۱ + \frac{\alpha P_x}{P_{n_۱}} \right). \qquad (۴۹.۸)$$

بنابراین گیرندهٔ ۱ به طور شایسته‌ای قادر به دریافت $\mathbf{x}_۱$ و $\mathbf{x}_۲$ می‌باشد.

این ایجاب می‌کند که نرخهای

$$R_۱ = W \log\left\{ ۱ + \frac{(۱-\alpha P_x)}{\alpha P_x + P_{n_۲}} \right\} + W \log\left(۱ + \frac{\alpha P_x}{P_{n_۱}} \right), \qquad (۵۰.۸)$$

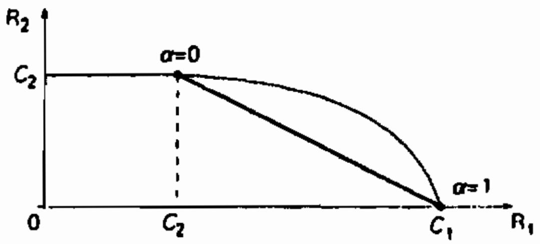
$$R_۲ = W \log\left\{ ۱ + \frac{(۱-\alpha) P_x}{\alpha P_x + P_{n_۲}} \right\}. \qquad (۵۱.۸)$$

را بتوان به دست آورد.

این نرخها بهتر از آنهایی هستند که با تقسیم زمان به دست می‌آوریم، همان‌طور که در شکل (۸. ۱۰) شرح داده شده است.

آنچه که واقعاً در این جا اتفاق افتاد به این طریق تشریح می‌شود. نوفهٔ بالای گیرندهٔ ۲ نسخهٔ تباهیده‌ای از $\mathbf{x}$ را به دست می‌آورد، در حالی که نوفهٔ پایین گیرندهٔ ۱ نیز اطلاعات اضافی دریافت می‌کند که آن را قادر می‌سازد که هنوز هم $\mathbf{x}$ اصلی را به دست آورد. برای مثال، این وضعیت برای فرستنده‌های تلویزیون رخ می‌دهد که برنامه‌هایی در سطح وضوح عالی تلویزیون (*HDTV*) پخش می‌کنند. برنامه‌ها با دستگاههای تلویزیون معمولی به خوبی دستگاههای *HDTV* خاص با کیفیت بالای تصویری دریافت می‌شود. با به کار بردن اصطلاح یکسانی همانند بحث بالا، می‌توانیم بگوییم که دستگاههای تلویزیون

معمولی می‌توانند سیگنال $x_۲$ را برای تولید تصاویر تلویزیون معمولی رمزگشایی کنند. دستگاههای *HDTV* علاوه بر $x_۲$ همچنین قادرند $x_۱$ را کدگشایی کنند، بنابراین کیفیت تصویری بالاتری به دست می‌آورند.



شکل ۸. ۱۰- ناحیهٔ ظرفیت برای کانال پخش گاوسی با دو گیرنده

اکنون می‌کوشیم ناحیهٔ ظرفیت یک حالت کلّی‌تر کانال پخش ضعیف شده را توصیف کنیم. با بیان مجدّد نابرابری فانو شروع می‌کنیم. همان‌طور که قبلاً گفته شد برای پیامهای کدگشایی‌ای که از طریق کانال پخش ارسال شده‌اند حالت گسسته را بررسی خواهیم کرد. قضیهٔ حاصل نقش مهمی را در تجزیه و تحلیل کانال پخش ضعیف شده بازی می‌کند.

قضیهٔ ۴.۸

فرض می‌کنیم یک رمزگذاری داده شده که می‌تواند پیامهای U^N به طول N به واژه‌های کدگذاری شدهٔ X^n به طول n که ورودی کانال را تشکیل می‌دهد برگرداند. خروجی کانال با Y^n نشان داده می‌شود، در حالی که پیامهای V^N بعد از رمزگشایی به دست می‌آید.

برای احتمال خطا در خروجی کانال داریم

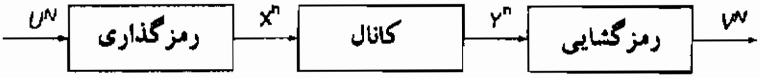
$$P_e \log(J-۱) + H(P_e) \geq \frac{۱}{N} H(X^n \big| Y^n), \qquad (۵۲.۸)$$

که در آن J اندازهٔ الفبای ورودی را نشان می‌دهد.

برهان

با به‌کارگیری نابرابری فانو (به قضیهٔ (۴.۴) مراجعه کنید)، برای نماد lام رمزی شدهٔ ورودی و نماد رمزگشایی شدهٔ خروجی می‌توان نوشت:

$$P_e \log(J-۱) + H(P_e) \geq H(U_l \big| V_l). \qquad (۵۳.۸)$$



شکل ۱۱.۸- طرح ارتباطی با کدگذاری

با محاسبهٔ میانگین هر دو طرف فرمول (۵۳.۸) برای تمام مقادیر l و با در نظر گرفتن این حقیقت که اندازه‌های اطلاع و احتمالهای خطا توابعی مقعّرند به دست می‌آوریم:

$$P_e \log(J-1) + H(P_e) \geq \frac{1}{N}\sum_{l=1}^{N} H(U_l | V_l). \qquad (54.8)$$

همچنین داریم

$$\begin{aligned} H(U^N | V^N) &= H(U_1 | V^N) + H(U_2 | V^N, U_1) + \cdots \\ &\quad + H(U_n | V^N, U_1, \ldots, U_{N-1}) \\ &\leq \sum_{l=1}^{N} H(U_l | V_l). \end{aligned} \qquad (55.8)$$

بنابراین، با کمک معادلهٔ (۵۴.۸) به دست می‌آوریم

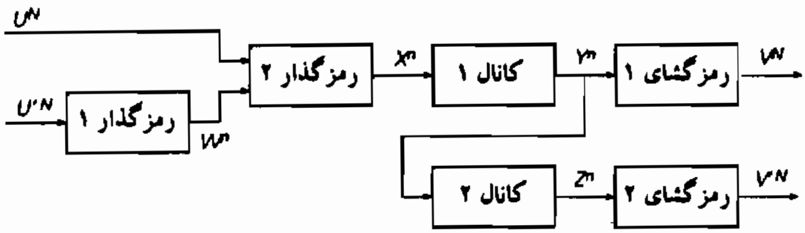
$$P_e \log(J-1) + H(P_e) \geq \frac{1}{N} H(U^N | V^N). \qquad (56.8)$$

سرانجام، چون یک رابطهٔ یک به یک بین U^N و X^n وجود دارد، می‌توان قضیهٔ (۱.۸) را به کار برد و نوشت

$$\begin{aligned} P_e \log(j-1) + H(P_e) &\geq \frac{1}{N} H(U^N | V^N) \\ &= \frac{1}{N}\{H(U^N) - I(U^N ; V^N)\} \\ &\geq \frac{1}{N}\{H(X^n) - I(X^n ; Y^n)\} \\ &= \frac{1}{N} H(X^n | Y^n), \end{aligned} \qquad (57.8)$$

به عبارتی که می‌خواستیم ثابت کنیم رسیدیم. ■

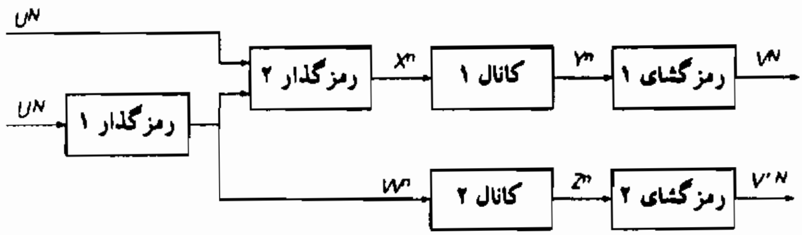
شکل (۱۲.۸) یک حالت توسعه یافتهٔ شکل (۷.۸) است، که برای تشریح قضیهٔ زیر به کار برده شده است.



شکل ۱۲.۸- کانال پخش تضعیف شده

در این جا فرض خواهیم کرد که سیگنالی که باید ارسال شود شامل وضعیت دو مؤلفه است: U^N و U'^N که در آن U'^N (و بنابراین W^n) اطلاعات ساخته شده برای هر دوگیرنده را نشان می‌دهد و در آن U^N اطلاعات اضافی برای گیرندهٔ ۱ را نشان می‌دهد. برای نمونه به مثال قبلی *HDTV* بیندیشید. نماد N طول پیامهای کدگذاری شده و کدگشایی‌شده را نشان می‌دهد؛ n طول کدواژه‌ها که ورودی و خروجی کانال را تشکیل می‌دهند می‌باشد.

اگر وضعیت را از نقطه نظر گیرندهٔ ۲ در نظر بگیریم، در این صورت دنبالهٔ رمزگذار ۲، کانال ۱ و کانال ۲ را می‌توان به عنوان یک کانال جدید تکی در نظر گرفت. این نتیجه در شکل (۱۳.۸) نشان داده شده است.



شکل ۱۳.۸- طرح دیگری از کانال پخش تضعیف شده

قضیهٔ ۵.۸

یک کانال پخش تضعیف شده را در نظر می‌گیریم. در صورتی که برای هر متغیّر تصادفی کمکی W و ورودی کانال X داریم

$$R_1 > I(X;Y|W) \tag{۵۸.۸}$$

یا

$$R_2 > I(W;Z), \tag{۵۹.۸}$$

در این صورت هیچ کدی با نرخهای (R_1, R_2) وجود نخواهد داشت که انتقال اطلاع را اجازه دهد بدون این که هیچ خطایی رخ دهد.

برهان

برای نرخ R_2 داریم

$$R_2 = \frac{1}{n} H(U'^N) = \frac{1}{n}\{I(U'^N; Z^n) + H(U'^N | Z^n)\}. \qquad (۶۰.۸)$$

عبارت زیر را می‌توان برای $I(U'^N; Z^n)$ به دست آورد

$$\begin{aligned} I(U'^N; Z^n) &= H(Z^n) - H(Z^n | U'^N) \\ &\le \sum_{l=1}^{n} H(Z_l) - \sum_{l=1}^{n} H(Z_l | U'^N, Z_{l-1}, \ldots, Z_1). \end{aligned} \qquad (۶۱.۸)$$

واضح است که برای تمام lها داریم

$$H(Z_l | U'^N, Z_{l-1}, \ldots, Z_1) \ge H(Z_l | U'^N, Z_{l-1}, \ldots, Z_1, Y_{l-1}, \ldots, Y_1)$$

و بنابراین

$$I(U'^N; Z^n) \le \sum_{l=1}^{n} \{H(Z_l) - H(Z_l | U'^N, Z_{l-1}, \ldots, Z_1, Y_{l-1}, \ldots, Y_1)\}. \qquad (۶۲.۸)$$

اگر U'^N و Y_{l-1}، ...، Y_1 معلوم باشند، در این صورت Z_1 به طور شرطی مستقل از Z_{l-1}، ...، Z_1 می‌باشد. در نتیجه

$$H(Z_l | U'^N, Z_{l-1}, \ldots, Z_1, Y_{l-1}, \ldots, Y_1) = H(Z_l | U'^N, Y_{l-1}, \ldots, Y_1)$$

و بنابراین

$$I(U'^N; Z^n) \le \sum_{l=1}^{n} I(Z_l; (U'^N, Y_{l-1}, \ldots, Y_1)). \qquad (۶۳.۸)$$

با جایگزینی $(U'^N, Y_{l-1}, \ldots, Y_1)$ توسط یک متغیّر تصادفی گسستهٔ جدید W_l این عبارت را می‌توان دوباره به فرم خیلی ساده‌تر زیر نوشت

$$\frac{1}{n} I(U'^N; Z^n) \le \frac{1}{n} \sum_{l=1}^{n} I(Z_l; W_l). \qquad (۶۴.۸)$$

سمت راست این عبارت را می‌توان به صورت بسط اطلاع متقابل $I(Z; W)$ بر نماد در نظر گرفت.

بنابراین، معادلهٔ (۶۴.۸) نتیجه می‌دهد

$$\frac{1}{n}I(U'^N;Z^n)\le I(Z;W).$$

با کمک معادلهٔ (۶۰.۸) برای R_2 می‌توان یافت:

$$\begin{aligned}R_2&=\frac{1}{n}\{I(U'^N;Z^n)+H(U'^N|Z^n)\}\\&\le I(Z;W)+\frac{1}{n}H(U'^N|Z^n).\end{aligned}\tag{۶۵.۸}$$

با معرّفی نابرابری فانو در این نقطه، مانند آنچه در قضیهٔ (۴.۸) داده شد، به دست می‌آوریم

$$\begin{aligned}P_e\log(J'-1)+H(P_e)&\ge\frac{1}{N}H(U'^N|Z^n)\\&\ge\frac{n}{N}(R_2-I(Z;W))\end{aligned}\tag{۶۶.۸}$$

که در آن J' اندازهٔ رمزگذار ورودی است.

تا زمانی که $R_2>I(Z;W)$، احتمال خطا نمی‌تواند به طور نامحدودی بیش از این نزول کند. بنابراین، قسمت دوم قضیه ثابت می‌شود.

قسمت اوّل قضیه را به همین منوال می‌توان ثابت کرد.

با یادآوری قضیهٔ (۱.۸) با نرخ R_1 می‌توان نوشت

$$\begin{aligned}R_1=\frac{1}{n}H(U^N|U'^N)&=\frac{1}{n}\{I(U^N;Y^n|U'^N)+H(U^N|Y^n,U'^N)\}\\&\le\frac{1}{n}\{I(X^n;Y^n|U'^N)+H(U^N|Y^n)\}.\end{aligned}\tag{۶۷.۸}$$

بنابر تعریف داریم

$$\begin{aligned}I(X^n;Y^n|U'^N)&=H(Y^n|U'^N)-H(Y^n|U'^N,X^n)\\&=\sum_{l=1}^{n}H(Y_l|U'^N,Y_{l-1},\dots,Y_1)\\&\quad-\sum_{l=1}^{n}H(Y_l|U'^N,X^n,Y_{l-1},\dots,Y_1).\end{aligned}\tag{۶۸.۸}$$

چون کانال بی‌حافظه است، Y_l نسبت به X^n تنها به X_l بستگی خواهد داشت و

بنابراین

$$H(Y_l \big| U'^n, X^n, Y_{l-1}, \ldots, Y_1) = H(Y_l \big| U'^N, X_l, Y_{l-1}, \ldots, Y_1).$$

علاوه بر این، با معادلهٔ (۶۸.۸) نتیجه می‌شود

$$\begin{aligned}\frac{1}{n} I(X^n; Y^n \big| U'^N) &= \frac{1}{n}\sum_{l=1}^{n} H(Y_l \big| U'^N, Y_{l-1}, \ldots, Y_1)\\ &\quad - \frac{1}{n}\sum_{l=1}^{n} H(Y_l \big| U'^N, X_l, Y_{l-1}, \ldots, Y_1)\\ &= \frac{1}{n}\sum_{l=1}^{n} I(X_l; Y_l \big| U'^N, Y_{l-1}, \ldots, Y_1)\\ &= I(X; Y \big| W). \qquad (۶۹.۸)\end{aligned}$$

مرحلهٔ آخر با معرّفی یک متغیّر تصادفی جدید W و با تشخیص جملهٔ شامل جمع‌بندی که می‌توان به عنوان یک بسط در نظر گرفت نتیجه می‌شود.

با ترکیب کردن این مشاهدات آخر، به دست می‌آوریم

$$\begin{aligned}H(U^N \big| Y^n) &\geq n\mathrm{R}_1 - I(X^n; Y^n \big| U'^N)\\ &= n\{\mathrm{R}_1 - I(X; Y \big| W)\}. \qquad (۷۰.۸)\end{aligned}$$

مجدّداً می‌توان نابرابری فانو را به کار برد، گرچه اکنون کران پایین برحسب $H(U^N \big| Y^n)$ داده شده است.

اگر احتمال خطا می‌بایستی به طور دلخواه کوچک انتخاب شود، در این صورت هرگز نباید $\mathrm{R}_1 > I(X; Y \big| W)$.

تنها ممکن است یک کد مناسب به دست آورد که اجازه دهد اطلاع با احتمال خطای کوچک دلخواهی ارسال شود، اگر هر دو شرط بیان شده توسط قضیه برقرار باشد. ■

قضیهٔ عکس را نیز می‌توان ثابت کرد. با وجود این، یک اثبات کامل ارائه نخواهیم کرد، چون قضیه با توجّه به قضیهٔ (۵.۸) بدیهی می‌شود.

قضیهٔ ۶.۸

اگر توزیع احتمال توأم ورودی X و متغیّر تصادفی کمکی W به قسمی باشد که

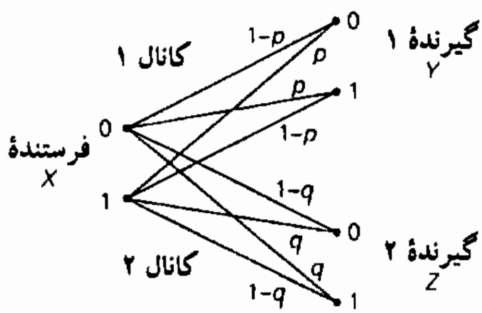
$$\mathrm{R}_1 \leq I(X; Y \big| W) \qquad (۷۱.۸)$$

و

$$\mathrm{R}_۲ \le I(W;Z), \qquad (۷۲.۸)$$

در این صورت به طور نظری کدی با نرخ $(\mathrm{R}_۱,\mathrm{R}_۲)$ می‌توان یافت که برای احتمال خطای کوچک دلخواهی مجاز باشد. ■

واضح است که ناحیهٔ ظرفیت کانالهای پخش تضعیف‌شدهٔ بی‌حافظه شامل غلاف محدّبی برای همهٔ $\mathrm{R}_۱$ و $\mathrm{R}_۲$ می‌باشد که شرایط بیان شده در قضیهٔ قبل را برای هر توزیع توأم W و Xای برقرار می‌کند.



شکل ۱۴.۸- کانال پخش با دو گیرنده

اکنون مثالی برای ظرفیت کانالهای دودویی متقارن فراهم می‌کنیم.

شکل (۱۲.۸) این حقیقت را نشان می‌دهد که کانال گیرندهٔ ۲ را می‌توان به صورت اتّصال سری کانال گیرندهٔ ۱ و کانال دیگری که با همدیگر کاملاً معادل کانال اصلی برای گیرندهٔ ۲ می‌باشند در نظر گرفت. کانال ۲ شکل (۱۴.۸) را می‌توان به عنوان کانال ۱ که با کانال دیگری دنبال می‌شود همچنان که در شکل (۱۵.۸) نشان داده شده است، در نظر گرفت.

بدون هیچ مشکلی می‌توان تأیید کرد که احتمالهای انتقال این توالی به ترتیب برابر q و $q-۱$ می‌باشند.

برای این که ظرفیت کانال را براساس قضیهٔ (۶.۸) تعیین کنیم یک متغیّر تصادفی کمکی W را معرّفی خواهیم کرد که توسط یک کانال متقارن دودویی با احتمال انتقال s به X متّصل شده است (شکل (۱۶.۸- الف) را ببینید).

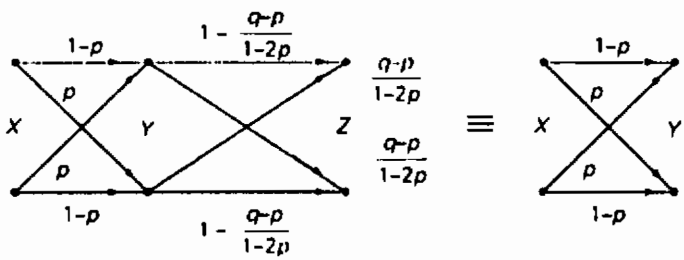
بنابر قضیهٔ (۶.۸) علاقه‌مند به یافتن $I(W;Z)$ و $I(X;(Y|W))$ هستیم. فرض کنید احتمالهای مربوط به W به ترتیب با α و $\alpha-۱$ نشان داده می‌شود. وقتی تقارن اتصال داخلی

را در نظر بگیریم روشن می‌شود که نرخها با $\alpha=\frac{۱}{۲}$ ماکسیمم می‌شوند. بنابراین بـا کمـک شکل (۱۶.۸- ب) به دست می‌آوریم

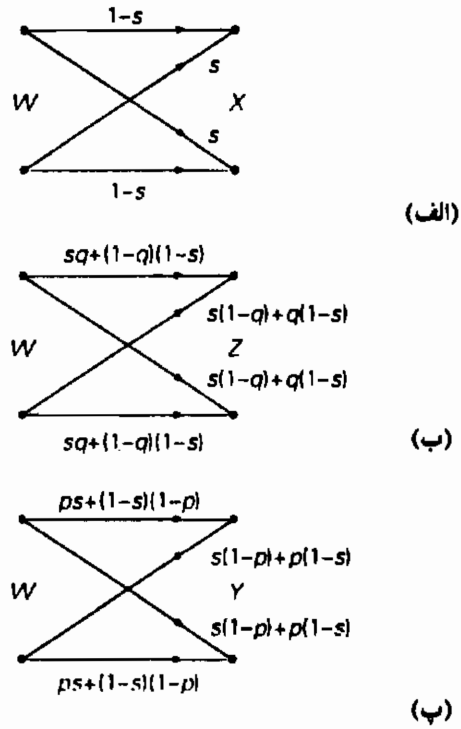
$$I(W;Z)=H(Z)-H(Z|W)$$

$$=۱+t\log t+(۱-t)\log(۱-t), \qquad (۷۳.۸)$$

با $t=sq+(۱-q)(۱-s)$.



شکل ۱۵.۸



شکل ۱۶.۸

برای $I(X;Y|W)$ با به کار بردن شکل (۱۶.۸- پ) می‌توان نوشت

$$I(X;Y|W) = H(Y|W) - H(Y|W,X)$$
$$= H(Y|W) - H(Y|X)$$
$$= -v\log v - (1-v)\log(1-v) + p\log p$$
$$+ (1-p)\log(1-p) \qquad (۷۴.۸)$$

با $v = sp + (1-p)(1-s)$.

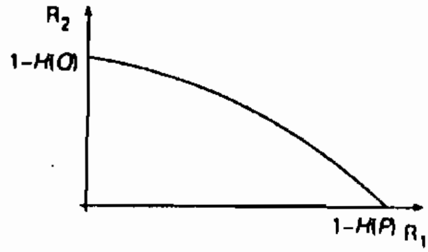
با تغییر مقدار s می‌توان مقادیری که کرانهای ناحیهٔ ظرفیت را تعریف می‌کنند محاسبه کرد. برای مثال، برای $s = 0$ نتیجه می‌شود $R_1 = 0$، چون $I(X;(Y|W)) = 0$ و

$$R_2 = 1 + q\log q + (1-q)\log(1-q) = 1 - H(Q).$$

وقتی $s = \frac{1}{2}$، نتیجهٔ محاسبه عبارت است از: $R_2 = 0$ و

$$R_1 = 1 + p\log p + (1-p)\log(1-p) = 1 - H(P).$$

ناحیهٔ ظرفیت در شکل (۱۷.۸) داده شده است.



شکل ۱۷.۸- ناحیهٔ ظرفیت کانال پخش متقارن دودویی

۴.۸ کانالهای دو-طرفه

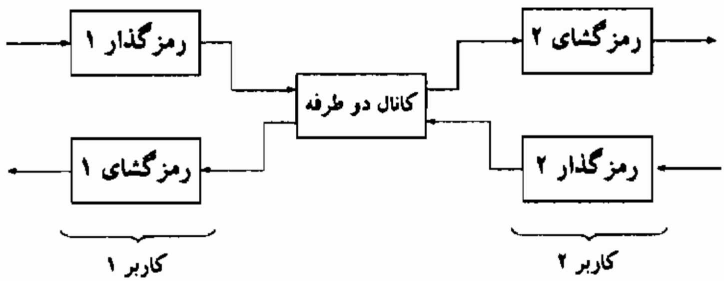
یک کانال دو-طرفه همواره شامل حداّقل دو بخش خواهد بود که هر دوی آنها می‌توانند به عنوان فرستنده و گیرنده عمل کنند. بدیهی است، اگر هر بخش نقشی تکی به عنوان فرستنده یا گیرنده اجرا کند با تعویض مکرّر نقش بخشها، این وضعیت رخ می‌دهد. با وجود این، عادی است که یک کانال دوطرفه هر دو بخش آن پیامهایشان را هم‌زمان به یک دیگر ارسال کنند و این که یک پیام در یک جهت با پیام در جهت دیگر تداخل کند.

شکل (۱۸.۸) یک نمودار کلّی از کانال دو-طرفه را نشان می‌دهد.

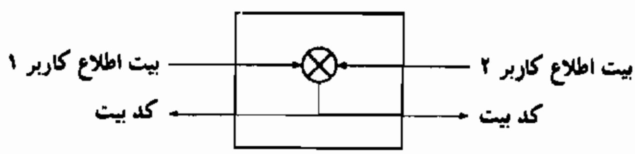
یکی از مشهورترین مثالهای کانال دو-طرفه ***کانال چندگانهٔ دودویی*** با الفبای دودویی ورودی و خروجی می‌باشد. با هر ضربهٔ (تیک‌تاک) ساعت هر کاربری بیت خودش از اطلاع را از طریق کانال ارسال می‌کند و هم‌زمان یک کد بیت دریافت می‌کند که در واقع «وَ» منطقی بیتهای ارسال شده است. اگر بیت اطلاع ارسال شده توسط یک کاربر، برابر ۱ باشد و بیت دریافت شده از کانال نیز ۱ باشد، در این صورت کاربر نتیجه می‌گیرد که بخش دیگر نیز بایستی ۱ بیت ارسال کرده باشد.

با وجود این، اگر بیت دریافت شده ۰ باشد در این صورت کاربر دیگر بایستی ۰ فرستاده باشد. اگر اولّین کاربر به جای یک ۰ بفرستد در این صورت قادر نخواهد بود که تعیین کند که آیا کاربر دیگر ۰ یا ۱ فرستاده است. گرچه، خوش‌بختانه کد هاگلبرگر[1] می‌تواند راه حلی برای این مسأله پیشنهاد کند (شکل (۱۹.۸) را ببینید).

هر رمزگذار می‌تواند یکی از دو وضعیت ۱ یا ۲ را اختیار کند. وضعیت جاری یک رمزگذار بستگی به آخرین کد بیت دریافتی و وضعیت زمان ضربهٔ قبلی ساعت خواهد داشت. بنابراین، هر دو رمزگذارها همواره در وضعیت یکسانی خواهند بود. یک رمزگذار تنها می‌تواند وارد وضعیت ۱ شود اگر بیت دریافتی برابر ۰ باشد یا وضعیت قبلی وضعیت ۲



شکل ۱۸.۸- کانال دوطرفه



شکل ۱۹.۸- کد هاگلبرگر

1. Hagelberger

باشد. در نتیجه، یک رمزگذار تنها می‌تواند داخل وضعیت ۲ شود اگر آخرین بیت دریافتی برابر ۱ باشد و رمزگذار قبلاً در وضعیت ۲ نباشد. در وضعیت ۱ رمزگذارها بیت دادهٔ بعدی را ارسال خواهند کرد. در وضعیت ۲ آنها مکمّل بیت دادهٔ قبلی خود را ارسال خواهند کرد.

شکل (۸. ۲۰) کد بیتهای حاصل را به عنوان تابعی از بیت داده‌ها نشان می‌دهد.

تعداد کد بیت لازم برای انتقال یک بیت سیگنال بستگی به بیت ارسال شده توسط هر دوی کاربرها دارد و برابر ۱ یا ۲ می‌باشد. اگر احتمال بیت ارسال شده توسّط کاربرها با $p(1) = p$ داده شده باشد، آن گاه برای هر بیت داده تعداد کلّ بیت کد برابر خواهد بود با

$$
\begin{aligned}
L &= 1 \times p^2 + 2p(1-p) + 2(1-p)^2 \\
&= 2 - p^2. \qquad (75.8)
\end{aligned}
$$

بنابراین، نرخ کد در هر جهت برابر خواهد شد با

$$
R = \frac{H(P)}{2 - p^2}, \qquad (76.8)
$$

که در آن $H(P) = -p\log p - (1-p)\log(1-p)$.

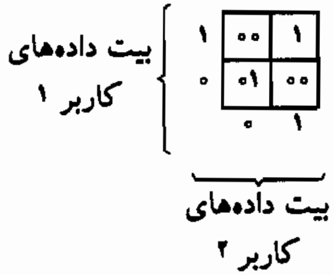
اگر $p = \frac{1}{2}$، آن گاه داریم

$$
R = \frac{1}{7/4} \approx 0.571.
$$

با وجود این، نرخ ماکسیمم به ازای $p = 0.63$ به دست می‌آید. این نرخ برابر است با $R = 0.593$. بنابراین، اگر کاربر بخواهد اطلاعی ارسال کند که برای آن $p = 0.5$ باشد، دراین صورت برای این که به نرخ ارسال ماکسیمم دست یابد، قبل از این که اطلاع به رمزگذار عبور دهد، ابتدا باید کدگذاری شود برای این که مطمئن باشد که احتمال ۱ برابر ۰٫۶۳ است. این کار را می‌توان با کدگشای هافمن انجام داد. چون یک کد هافمن برای انتقال دادهٔ اصلی با احتمال متفاوت بر نماد به کار برده می‌شود با کدواژه‌ها با نمادهای هم‌احتمال، بدیهی است معادلهٔ وارون به اثر مطلوب نایل خواهد شد؛ نمادها از نمادهای هم‌احتمال به نمادهای با احتمال متفاوت انتقال داده خواهد شد.

در این صورت، گیرنده کد هافمن را به کار خواهد برد. علاوه بر کد هاگلبرگر، کد شالک‌ویک[1] را نیز می‌توان به کار برد، که نرخ ۰٫۶۱۹ بیت بر ثانیه به طور هم‌زمان در هر دو جهت خواهد داد.

1. Schalkwijk



شکل ۸. ۲۰

امکان‌پذیر نیست که یک عبارت کلّی برای ظرفیت کانال دوطرفهٔ بی‌حافظه به دست آورد. با وجود این، هنوز هم می‌توان مقادیر کرانهای داخلی و خارجی ظرفیت را محاسبه کرد.

فرض کنید X و X' متغیّرهای تصادفی روی الفبای ورودی کانال باشند و Y و Y' متغیّرهای تصادفی روی الفبای خروجی کانال باشند.

قضیهٔ زیر یک کرانهٔ خارجی برای ناحیهٔ ظرفیت فراهم می‌کند.

قضیهٔ ۷.۸

فرض کنید (X,X',Y,Y') دارای توزیعهای احتمال توأم $p(x_l,x_l')$ و $p(y_l,y_l'|x_l,x_l')$ برای هر توزیع احتمال توأم $p(x_l,x_l')$ باشد. در این صورت ناحیهٔ ظرفیت کانال دوطرفهٔ بی‌حافظه (شکل (۲۱.۸) را ببینید) به کرانهایی محصور خواهد بود که برای آن

$$(\mathrm{R}_1,\mathrm{R}_2): \mathrm{R}_1 \le I(X;Y'|X') \text{ و } \mathrm{R}_2 \le I(X';Y|X). \qquad (۷۷.۸)$$

برهان

اگر برای زوج $(\mathrm{R}_1,\mathrm{R}_2)$ شرایط قضیه برقرار نباشد باید ثابت کنیم که امکان ندارد کدهایی بیابیم که اجازه دهند احتمال خطا را به مقدار کوچک دلخواهی تقلیل داد. بدیهی است که داریم

$$H(U^N) = I(U^N;Y'^n) + H(U^N|Y'^n) \qquad (۷۸.۸)$$

و

$$H(U^N|U'^N) = I(U^N;Y'^n|U'^N) + H(U^N|Y'^n,U'^N). \qquad (۷۹.۸)$$

چون یک تناظر یک به یک از یک طرف بین U^N و X^n و از طرف دیگر

بین U'^N و X'^n وجود دارد، داریم

$$H(U^N \mid U'^N) = I(X^n;Y'^n \mid X'^n) + H(X^n \mid Y'^n, X'^n).$$

علاوه بر این، داریم

$$I(X^n;Y'^n \mid X'^n) \le \sum_{l=1}^{n} I(x_l;y'_l \mid X'^n)$$

$$= \sum_{l=1}^{n}\left\{\sum_{x_l, X'^n, y'_l} p(x_l, X'^n, y'_l)\log\left\{\frac{p(x_l \mid N'^n, y'_l)}{p(x_l \mid X'^n)}\right\}\right\}$$

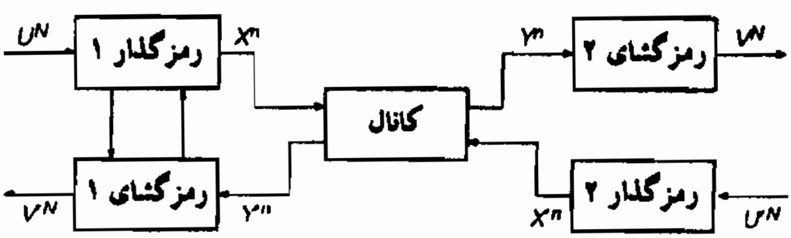
$$= \sum_{l=1}^{n}\left\{\sum_{x_l, X'^n, y'_l} p(x_l, X'^n, y'_l)\log\left\{\frac{\frac{p(x_l, X'^n, y'_l)}{p(X'^n, y'_l)}}{\frac{p(x_l, X'^n)}{p(X'^n)}}\right\}\right\}$$

$$= \sum_{l=1}^{n}\left\{\sum_{x_l, X'^n, y'_l} p(x_l, X'^n, y'_l)\log\left\{\frac{p(y'_l \mid x_l, X'^n)p(x_l, X'^n)}{\frac{p(y'_l \mid X'^n)p(X'^n)p(x_l, X'^n)}{p(X'^n)}}\right\}\right\}$$

$$= \sum_{l=1}^{n}\{H(Y'_l \mid X'^n) - H(Y'_l \mid X_l, X'^n)\}$$

$$\le \sum_{l=1}^{n} H(Y_l \mid X'_l) - \sum_{l=1}^{n} H(Y'_l \mid (X_l, X'_l, X'_{l-1}, , X'_1,)); \qquad (۸۰.۸)$$

چون یک کانال بی‌حافظه داریم:



شکل ۲۱.۸- کانال دو-طرفه

$$H(Y'_l|(X_l, X'_l, X'_{l-1}, \ldots, X'_1)) = H(Y'_l| X_l, X'_l),$$

و بنابراین، از معادلهٔ (۸. ۸۰) به دست می‌آوریم

$$I(X^n; Y'^n| X'^n) \le \sum_{l=1}^{n} H(Y'_l| X'_l) - \sum_{l=1}^{n} H(Y'_l| X_l, X'_l)$$

$$= \sum_{l=1}^{n} I(X_l; Y'_l| X'^n)$$

$$= nI(X; Y'| X'). \tag{۸۱.۸}$$

برای نرخ هر نماد R_1، داریم:

$$\mathrm{R}_1 = \frac{1}{n} H(U^N| U'^N) = \frac{1}{n}\{I(X^N; Y'^n| X'^n) + H(X^n| Y'^n, X'^n)\}$$

$$\le I(X| Y', X') + \frac{1}{n} H(X^n| Y'^n, X'^n) \tag{۸۲.۸}$$

و بنابراین

$$H(X^n| Y'^n, X'^n) \ge n\{\mathrm{R}_1 - I(X; Y'| X')\}. \tag{۸۳.۸}$$

با کمک نابرابری فانو باید داشته باشیم

$$n\{P_e \log(J-1) + H(P_e)\} \ge H(X^n| Y'^n, X'^n) \ge n\{\mathrm{R}_1 - I(X; Y'| X')\}. \tag{۸۴.۸}$$

اگر احتمال خطا به صفر میل کند، در این صورت داریم

$$\mathrm{R}_1 \le I(X; Y'| X'). \tag{۸۵.۸}$$

چون نقش X و X' در این عبارتها متقارن هستند، روشن است که برای R_2 باید داشته باشیم

$$\mathrm{R}_2 \le I(X'; Y| X).$$

این همان عبارتی است که می‌خواستیم ثابت کنیم. ■

در قضیهٔ بالا فرض کردیم که رمزگذارها در حدّی قابلیّت همکاری با یکدیگر را دارند. این فرض به صراحت بیان نشد، ولی با به کارگیری توزیعهای احتمال توأم $p(x_l, x'_l)$ به جای $p(x_l) \cdot p(x'_l)$ از آن استفاده شد.

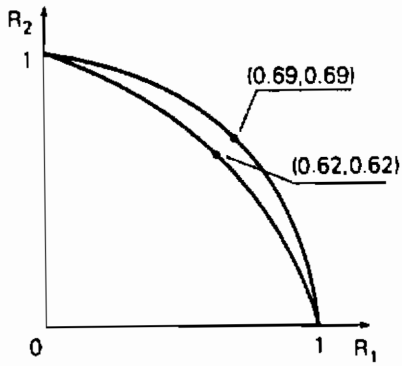
با وجود این، در عمل، یک رمزگذار تنها اطلاعاتی جزئی از کدواژه‌ای که رمزگذار

دیگر می‌خواهد بفرستد دارد. به محض دریافت اوّلین نماد، عدم‌حتمیت بیشترین خواهد بود، ولی در جهت پایان بلوک ارسال شده، حتمیت نسبت به پیام افزایش خواهد یافت. این بدین علّت است که کران داده شده توسط قضیه یک کران بیرونی است. یک کران درونی با این فرض که رمزگذارها کاملاً مستقل از یکدیگر عمل می‌کنند می‌توان به دست آورد: $p(x_1, x_1') = p(x_1) \cdot p(x_1')$. برای همهٔ (X, X', Y, Y') با توزیع احتمال توأم متناظر $p(x_1)p(x_1')p(y_1', y_1'| x_1, x_1')$ این مطلب که غلاف محدّب مجموعهٔ

$$\{(\mathrm{R}_1, \mathrm{R}_2) \mid \mathrm{R}_1 \le I((X \mid X'); Y) \text{ و } \mathrm{R}_2 \le I((X \mid X'); Y)\}$$

در ناحیهٔ ظرفیت کانال دوطرفه محصور شده است معتبر می‌باشد.

این کرانهای درونی و بیرونی در شکل (۲۲.۸) داده شده است.



شکل ۲۲.۸- کرانهای درونی و بیرونی برای کانال دو-طرفه

۵.۸ تمرینها

۱.۸ یک کانال چندمدخلی را با دو فرستنده و با ورودیهای دودویی در نظر بگیرید. خروجی به صورت زیر داده شده است

$$Y = X_1 * X_2$$

که در آن * عمل تعریف نشده‌ای را نشان می‌دهد.

(الف) اگر عمل * یک ضرب را نشان دهد، یک کانال ضربی دودویی (BMC) خواهیم داشت. ناحیهٔ ظرفیت را برای این کانال تعیین کنید.

(ب) کانال چند-مدخلی پاک‌شدگی دودویی ($BEMC$) نتیجه خواهد شد در حالتی که

در آن عمل * جمع را نشان دهد. خروجی فرض شده است کـــه سه‌ســه‌ای باشــد. ناحیهٔ ظرفیت را برای کانال ضربی دودویی به دست آورید.

۲.۸ یک کانال چند-مدخلی را با k کاربر که در آن نوفهٔ گاوسی جمعی موجب تداخل می‌شود در نظر بگیرید. فرض کنید که هر کاربر در سطح توان سیگنال یکسان P_x ارسال می‌کند. در این صورت برای هر کاربر نسبت سیگنال به نوفه برابــر $\frac{P_x}{P_n}=10$ می‌باشد.

(الف) ظرفیت کانال را برای حالتی که در آن تقسیم فرکانس به کار برده شده و به هـــر یک از k کاربر یک پهنای باند مساوی w نسبت داده شده است محاسبه کنید.

(ب) ظرفیت را در واحد پهنای باند برای حالت شرح داده شده در (الف) به عنوان تابعی از k رسم کنید.

(پ) عبارتی برای ظرفیت موجود برای هر کاربر به دست آورید، اگر تقسیم زمــان بــه جای تقسیم فرکانس به کار برده شود، فرض کنید که هر کاربر ممکن است برای یک مدّت زمان یکسان به کانال دسترسی داشته باشد.

۳.۸ دو فرستنده می‌توانند به کانال چند-مدخلی گاوسی دسترسی پیدا کنند. فرســتندهٔ ۱ نسبت سیگنال به نوفهٔ ۲۰ و فرستندهٔ ۲ نسبت سیگنال به نوفهٔ ۱۰ دارد.

(الف) ظرفیت برحسب بیت بر ارسال فرستندهٔ ۱ را با این فــرض کــه فرستندهٔ ۱ دارای دسترسی خاصی به کانال است محاسبه کنید. این محاسبه را برای فرستندهٔ ۲ تکــرار کنید.

(ب) ناحیهٔ ظرفیت کانال چند-مدخلی گاوسی را برای حالتی که در آن هر دو فرســتنده هم‌زمان از کانال استفاده می‌کنند به دست آورید.

۴.۸ کانال پخشی را با دو گیرنده در نظر بگیرید. کانال بیــن فرســتنده و گــیرندهٔ ۱ را می‌توانید به عنوان یک کانال متقارن دودویی در نظر بگیرید که برای آن احتمــال ارسال نادرست برابر است با $p=0{,}1$. برای کانال بین فرستنده و گیرندهٔ ۲ احتمـــال ارسال نادرست برابر است با $q=0{,}6$.

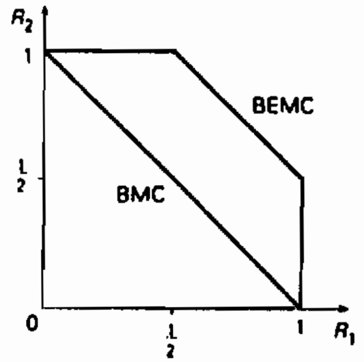
(الف) ظرفیت کانال بین فرستنده و گیرندهٔ ۱ را محاسبه کنید. همچنین ظرفیت کانال بین فرستنده و گیرندهٔ ۲ را به دست آورید.

(ب) دو عبارت برحسب اطلاع متقابل که با یکدیگر ناحیهٔ ظرفیت را برای کانال پخش شرح داده شده در این جا تعریف می‌کنند بیابید.

(پ) ناحیهٔ ظرفیت این کانال پخش را رسم کنید.

۶.۸ جوابها

۱.۸ (الف) اگر $X_1 = 1$، در این صورت فرستندهٔ ۲ قادر خواهد بود اطلاع را به گیرنده با نرخ ماکسیمم ۱ بیت در هر ارسال بفرستد. همین مطلب برای فرستندهٔ ۱ برقرار می‌باشد. وقتی $X_2 = 1$، فرستندهٔ ۱ نیز اطلاع را با نرخ ۱ بیت به گیرنده می‌فرستد. بنابراین، دو نقطه به ترتیب روی محورهای R_2 و R_1 تعریف شده است. تقسیم زمان میسّر می‌سازد که هر ترکیبی از نرخها که برای آن $R_1 + R_2 = 1$ به دست آورد. ناحیهٔ ظرفیت متناظر در شکل (۲۳.۸) داده شده است.



شکل ۲۳.۸- ناحیه‌های ظرفیت برای *BMC* و *BEMC*

(ب) اگر خروجی Y مقدار ۰ یا ۲ را اختیار کند، در این صورت هیچ گونه عدم‌حتمیت با بیت اطلاع فرستاده شده توسط دو فرستنده وجود ندارد. با وجود این، وقتی Y برابر ۱ است عدم‌حتمیت وجود دارد.

اگر $X_1 = 0$، آن گاه فرستندهٔ ۲ می‌تواند اطلاع به گیرنده با نرخ ۱ بیت بر ارسال بفرستد: $R_2 = 1$. یعنی، R_1 برابر ۱ خواهد بود وقتی $X_2 = 0$. بنابراین، دو نقطهٔ کران بیرونی ناحیهٔ ظرفیت تعریف شده‌اند. اگر $R_1 = 1$، آن گاه بیتهای فرستاده شده توسط فرستندهٔ ۱ را می‌توان به عنوان نوفه توسط فرستندهٔ ۲ در نظر گرفت. نصف بیتها از فرستندهٔ ۱ که به بیت اطلاع فرستندهٔ ۲ اضافه شده است صفر خواهد

بود، در حالی که نصف دیگر بیتها یک خواهد بود. برای فرستندهٔ ۲، کانال چند-مدخلی را می‌توان به روشی مانند کانال پاک‌شدگی دودویی (BEC) با احتمال انتقال (۱,۲) شرح داد. در تمرین (۷.۴) قبلاً پیدا کردیم که ظرفیت کانال پاک‌شدگی دودویی برابر $C=۱-p$ می‌باشد. بنابراین، در این جا $C=\frac{۱}{۲}$ به دست خواهیم آورد.

یعنی وقتی فرستندهٔ ۱ در حال ارسال اطلاع با نرخ $R_1=۱$ است، فرستندهٔ ۲ هنوز هم قادر خواهد بود $\frac{۱}{۲}$ بیت بر ارسال به گیرنده بفرستد. بالعکس، اگر برای فرستندهٔ ۲، $R_2=۱$، در این صورت فرستندهٔ ۱ نیز قادر خواهد بود $\frac{۱}{۲}$ بیت ارسال کند. ناحیهٔ ظرفیت حاصل در شکل (۲۳.۸) داده شده است.

۲.۸ (الف) در حالت تقسیم فرکانس چند-مدخلی داریم

$$C^* = \frac{W}{k}\log\left\{۱+\frac{kP_x}{P_n}\right\} = \frac{W}{k}\log\{۱+۱۰k\}.$$

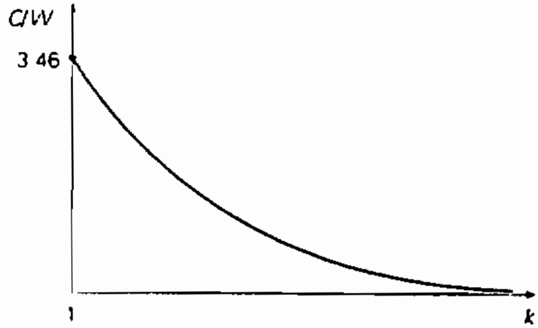
(ب) جدول زیر مقادیر $\frac{C}{W}$ متناظر با چندین مقدار k را می‌دهد

k	$\frac{C}{W}$
۱	۳٫۴۶
۲	۲٫۲۰
۳	۱٫۶۵
۴	۱٫۴۰
۵	۱٫۱۳
۱۰	۰٫۶۶
۲۰	۰٫۳۸
۳۰	۰٫۲۷
۴۰	۰٫۲۲
۵۰	۰٫۱۸
۱۰۰	۰٫۱۳

برای مقادیر بزرگ k نتیجه می‌شود که $\frac{C}{W}\approx\frac{۱}{k}\log ۱۰k$. منحنی متناظر با $\frac{C}{W}$ در شکل (۲۴.۸) رسم شده است.

(پ) چون تقسیم زمان به طور مؤثری نتیجهٔ ظرفیت یکسانی مانند تقسیم فرکانس

می‌دهد، نتایج یکسانی برای (الف) و (ب) به دست خواهیم آورد.



شکل ۲۴.۸- ظرفیت کانال چند-مدخلی تمرین (۲.۸)

۳.۸ ظرفیت به صورت زیر داده شده است

$$C = W \log\left(1 + \frac{P_x}{P_n}\right) \text{ ثانیه / بیت}$$

یا

$$C = \frac{1}{2} \log\left(1 + \frac{P_x}{P_n}\right) \text{ ارسال / بیت.}$$

بنابراین، برای فرستندهٔ ۱ و ۲ به ترتیب نتیجه می‌شود که

$$C_1 = \frac{1}{2} \log\left(1 + \frac{P_1}{P_n}\right) = \frac{1}{2} \log 21 = 2{,}20$$

و

$$C_2 = \frac{1}{2} \log\left(1 + \frac{P_2}{P_n}\right) = \frac{1}{2} \log 11 = 1{,}73,$$

(ب) دو مقدار محاسبه شده برای جواب قبلی (الف) متناظر با دو کران ناحیهٔ ظرفیت می‌باشد.

نظریه بیان می‌کند که وقتی فرستندهٔ ۲ اطلاع را با نرخ ماکسیمم ارسال کند، نرخ ماکسیمم برای فرستندهٔ ۱ برابر است با $\frac{1}{2} \log\left(1 + \frac{P_1}{P_2 + P_n}\right)$. با این وصف سیگنال ارسال‌شده توسّط فرستندهٔ ۲ را ممکن است توسّط فرستندهٔ ۱ به عنوان یک

منبع نوفهٔ اضافی در نظر گرفت.

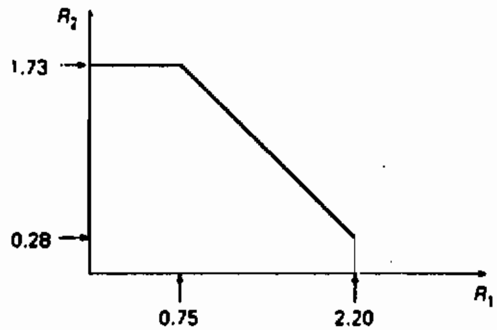
چون $P_1 = ۲۰ P_n$ و $P_۲ = ۱۰ P_n$ نتیجه می‌شود که $\frac{P_1}{P_۲ + P_n} = \frac{۲۰ P_n}{۱۱P_n} = \frac{۲۰}{۱۱}$ و بنابراین

$$\frac{1}{2}\log\left(1+\frac{P_1}{P_2+P_n}\right) = \frac{1}{2}\log\left(1+\frac{20}{11}\right) = \frac{1}{2}\log\frac{31}{11} = 0{,}75.$$

به همین روش می‌توان نتیجه گرفت که وقتی فرستندهٔ ۱ اطلاع با نرخ ماکسیمم می‌فرستد، نرخ ماکسیمم برای فرستندهٔ ۲ برابر است با

$$\frac{1}{2}\log\left(1+\frac{P_2}{P_1+P_n}\right) = \frac{1}{2}\log\left(1+\frac{10}{21}\right) = \frac{1}{2}\log\frac{31}{21} = 0{,}28.$$

بنابراین، تمام ناحیهٔ ظرفیت کامل با شکل (۲۵.۸) داده شده است.



شکل ۲۵.۸- ناحیهٔ ظرفیت کانال چند-مدخلی گاوسی تمرین (۳.۸)

۴.۸ (الف) چون با دو کانال متقارن دودویی سروکار داریم می‌توانیم بـه مثـال (۳.۴) مراجعه کنیم، که در آن قبلاً ظرفیت برای حالت کلّی محاسبه شده است.

برای گیرندهٔ ۱ داریم

$$C_1 = 1 - H(P) = 1 + 0{,}1\log 0{,}1 + 0{,}9\log 0{,}9 = 0{,}531.$$

ظرفیت برای گیرندهٔ ۲ به صورت زیر است

$$C_2 = 1 - H(Q) = 1 + 0{,}6\log 0{,}6 + 0{,}4\log 0{,}4 = 0{,}029.$$

(ب) کانال برای گیرندهٔ ۲ به عنوان دو کانال متوالی در نظر گرفته شـده اسـت، یعنـی کانال برای گیرندهٔ ۱ و یک کانال اضافی، که تضمین می‌کنـد کـه توالـی دارای

احتمالهای انتقال یکسانی به صورت کانال اصلی است.

ناحیهٔ ظرفیت به طور ضمنی با $I(W;Z)$ و $I(X;(Y|W))$ داده شده است که در آن X ورودی کانال تولید شده توسط فرستنده را نشان می‌دهد و Y و Z خروجیهای کانال به ترتیب برای فرستندهٔ ۱ و فرستندهٔ ۲ را نشان می‌دهند. W یک متغیّر کمکی است.

عبارت برای $I(W;Z)$ عبارت است از

$$I(W;Z)=H(Z)-H(Z|W)=1+t\log t+(1-t)\log(1-t)$$

که برای آن

$$t=sq+(1-q)(1-s)=0{,}6s+0{,}4(1-s)=0{,}2s+0{,}4$$

و s برابر احتمال انتقال است که مقدارش تغییر می‌کند، با فرض این که کانال بین W و X متقارن دودویی است.

برای $I(X;(Y|W))$ می‌توان نوشت

$$\begin{aligned}I(X;(Y|W))&=-v\log v-(1-v)\log(1-v)\\&\quad+p\log p+(1-p)\log(1-p)\\&=-v\log v-(1-v)\log(1-v)\\&\quad+0{,}1\log 0{,}1+0{,}9\log 0{,}9\\&=-v\log v-(1-v)\log(1-v)-0{,}469\end{aligned}$$

با

$$v=sp+(1-p)(1-s)=0{,}1s+0{,}9(1-s)=0{,}9-0{,}8s.$$

(پ) هیچ عبارت صریحی برای ناحیهٔ ظرفیت وجود ندارد. با وجود این، نقاط ناحیه را با تغییر دادن مقدار s بین ۰ و ۱ می‌توان به دست آورد.

اگر $s=0$ نتیجه می‌شود که $t=0{,}4$ و $v=0{,}9$ و در نتیجه داریم

$$I(W;Z)=1+0{,}4\log 0{,}4+0{,}6\log 0{,}6=0{,}029,$$

$$I(X;(Y|W))=-0{,}9\log 0{,}9-0{,}1\log 0{,}1-0{,}47=0.$$

اگر $s=\frac{1}{2}$، نتیجه می‌شود که $t=0{,}5$ و $v=0{,}5$، بنابراین

$$I(W;Z)=1+0{,}5\log 0{,}5+0{,}5\log 0{,}5=0,$$

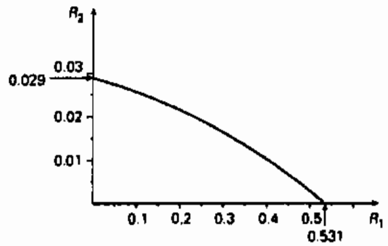
$$I(X;(Y|W)) = -0{,}5\log 0{,}5 - 0{,}5\log 0{,}5 - 0{,}47 = 1 - 0{,}47$$
$$= 0{,}531.$$

این کاملاً متناظر با مقادیر به دست آمده برای ظرفیت در (الف) است، به فرض این که فرستندهٔ دیگر از ارسال کردن خودداری کند.

برای $s = 0{,}1, 0{,}2, 0{,}3, 0{,}4$ می‌توان مقادیر زیر را محاسبه کرد:

$I(W;Z)$	$I(X;(Y\|W))$
۰٫۰۱۹	۰٫۲۱۰
۰٫۰۱۰	۰٫۳۵۷
۰٫۰۰۵	۰٫۴۵۵
۰٫۰۰۱	۰٫۵۱۲

شکل (۲۶.۸) را ببینید.



شکل ۲۶.۸- ناحیهٔ ظرفیت کانال پخش تمرین (۴.۸)

۹
کدهای تصحیح‌کنندهٔ خطا

۱.۹ مقدّمه

همچنان که در فصل چهارم به دست آوردیم، می‌توان در اصل ارسال بدون-خطـا از طریق یک کانال ارتباطی گسسته به دست آورد به شرط آن که مقدار اطلاعـی کـه بـاید منتقل شود حداّکثر C باشد. این قضیه هیچ کمکی در ساختار کدهای واقعـی نمی‌کنـد. در واقع تنها به وجود چنین کدی تأکید می‌کند.

در این فصل طرح تشخیص-خطا و تصحیح خطـای کدهـا را بـا جزئیـات بیشـتری بررسی خواهیم کرد. برای مطالعهٔ این کدها دو زمینهٔ جالب وجـود دارد، یعنـی سـاختمان کدها از یک طرف و رفتار آنها نسبت به تشخیص و تصحیح خطـا در اثنـای کدگشـایی واژه‌های دریافت شده از طرف دیگر. به خصوص کدگشایی اغلب در عمل پیچیده است و از این رو به توجّه بیشتری نیاز دارد. در این فصل به کدهای بلوکی خطی ساده برای کانال متقارن دودویی بی‌حافظه با احتمال خطای p اکتفا می‌کنیم (فصل چهارم را نیز ببینید).

دنبالهٔ نمادهای تولید شده توسط منبع اطلاع را به بلوکهای k نمادی تقسیم می‌کنیـم. پیام u شامل k نماد را با u_1، u_2، ...، u_k نمایش می‌دهیم. فرض می‌کنیم که این نمادهـا از الفبایی که با الفبای کانال یکسان است به وجود می‌آیند. بنابراین برای یک کانال دودویی داریم $u_i \in \{0,1\}$، $i = 1,\dots,k$. علاوه بر این فرض می‌کنیـم کـه $p(u_i = 0) = p(u_i = 1) = \frac{1}{2}$. یعنی، فرض می‌کنیم که از یک منبع کد اپتیمال بهره برده‌ایم. چون کانال متقـارن اسـت، همچنین این توزیع احتمال منبع اپتیمال است که دست‌یابی به ظرفیت کانال را مطمئن می‌سازد.

بنابراین تعداد پیامهای ممکن u برابر $M=2^k$ است، و هر پیام u دارای احتمال برابر $p(u)=2^{-k}$ می‌باشد.

وقتی یک کد کانال طرح می‌کنیم یک یا بیشتر از یک بررسی توازن در یک روش روشن به این پیامها اضافه می‌کنیم. طول M کدواژهٔ به دست آمده با این روش را با n نشان می‌دهیم. اکنون نرخ کد برابر است با

$$R=\frac{\log M}{n}$$

$$=\frac{\log 2^k}{n}=\frac{k}{n}\ \text{بیت / نماد}, \tag{1.9}$$

و می‌توان آن را به عنوان اندازه‌ای برای حشوی که اضافه شده است برای این‌که تشخیص یا تصحیح خطاها را میسّر سازد تفسیر نمود.

بنابراین ساختمان یک کد را می‌توان با انتخاب M کدواژهٔ کانال از کلّ 2^n کدواژهٔ ممکن تلقّی کرد. اکنون کدگشایی را می‌توان به صورت این تصمیم در نظر گرفت که کدام واژهٔ x کانال براساس دریافت واژهٔ y (احتمالاً دگرشکلی شده) ارسال شده است.

اگر واژهٔ y دریافت شده با یک کدواژه برابر نباشد در این صورت حدّاقل یک خطا رخ داده است. اگر y برابر با یک کدواژه باشد در این صورت ممکن است ارسالی بی‌خطا داشته باشیم، ولی این به هیچ وجه حتمی نیست.

می‌توان با بررسی $(n-k)$ برابری به حشو پیام افزود به قسمی که تشخیص یا تصحیح خطاها در کانال ارتباطی را میسّر سازد. با وجود این، اصلاح احتمال خطا همواره به بهای مقدار اطلاع ارسال شده به ازای هر نماد منبع، R، می‌باشد.

کدهای موسوم به ***کدهای تکراری*** را بررسی می‌کنیم. در این کدها یک نماد به تعدادی از دفعات تکرار می‌شود. فرض کنید یک منبع دو نماد ۰ و ۱ را تولید می‌کند. فرض کنید که احتمال خطا برابر $p=0.01$ باشد. در این حالت احتمال ارسال صحیح یک نماد 0.99 است. اگر رمزگذاری‌ای به کار بریم به طوری‌که کدواژه ۰۰۰ به نماد ۰ و کدواژه ۱۱۱ را به ۱ (هر نماد تولید شدهٔ منبع سه مرتبه تکرار می‌شود) نسبت دهد، در این صورت نرخ از $R=1$ به $R=\frac{1}{3}$ تنزل پیدا می‌کند. با وجود این، احتمال دریافت صحیح افزایش خواهد یافت. اگر کدواژه‌ها با خطاها متأثر نشوند و یا اگر درست یک نماد کدواژه تغییر کند دریافت صحیح رخ می‌دهد. در حالت اخیر براساس اکثریّت آرا نماد ورودی صحیح به دست خواهد آمد.

پیام بازسازی شده	پیامهای دریافت شده	کدواژه	پیام
۰	۰۰۰ ۱۰۰ ۰۱۰ ۰۰۱	۰۰۰	۰
۱	۱۱۰ ۱۰۱ ۰۱۱ ۱۱۱	۱۱۱	۱

اگر ۲ یا ۳ خطا در کدواژه وجود داشته باشد تصمیم غلط گرفته خواهد شد.
اکنون احتمال خطا، P_e، برابر می‌شود با

$$P_e = ۱ - \{(۰{,}۹۹)^۳ + ۳ \times ۰{,}۰۱(۰{,}۹۹)^۲\} \approx ۳ \times ۱۰^{-۴}.$$

به طور کلّی‌تر در یک کد $(n,۱)$-تکرار یک نماد $(n-۱)$ مرتبه تکرار می‌شود. از این رو تکرار نمادهای ورودی اثر سودمندی بر احتمال خطا دارد.
تصحیح خطا با شمارش صفرها در یک واژهٔ دریافت شده و با تصمیم رفتن بـه نمـاد صفر اگر این عدد بزرگتر از $\frac{n}{۲}$ باشد امکان‌پذیر می‌باشد. نرخ ایـن کـد برابـر بـا $R = \frac{۱}{n}$ می‌باشد. در جدول زیر نتیجهٔ این کد برای چندین مقدار از n برای احتمال خطـای $p = ۰{,}۰۱$ نشان داده شده است.

نرخ کد	باقی‌ماندهٔ احتمال خطا	طول n
۱	$۱۰^{-۲}$	۱
۱/۳	۳×۱۰^{-۴}	۳
۱/۵	$۱۰^{-۶}$	۵
۱/۷	۴×۱۰^{-۷}	۷
۱/۹	$۱۰^{-۸}$	۹
۱/۱۱	۵×۱۰^{-۱۰}	۱۱

این جدول نشان می‌دهد که چنین کدی ویژگی تصحیح-خطای خوبی دارد، ولی نرخ کد پایین است. اگر بگذاریم تعداد تکرارها و از این رو تعداد نمادها در هر پیام به بی‌نهایت میل کنند، احتمال خطا واقعاً به صفر میل می‌کند، ولی مقدار اطلاع انتقال یافته نیز این کد را در عمل نامطلوب می‌کند.

یک رهیافت صحیح را با ساختن کدهای (n,k) می‌توان به دست آورد که در آن هر دوی n و k افزایش یابند ولی تعداد بررسیهای توازن حفظ شود، یعنی $n-k$ محدود شده است. از این طریق می‌توان یک عامل توازن بین تعداد خطاهایی که باید تصحیح شود و نرخ کد مطلوب به دست آورد.

۲.۹ کدهای بلوکی خطی

همان‌طور که قبلاً ذکر شد در این جا خود را به کدهای بلوکی خطی محدود می‌کنیم. ***وزن*** $w(a)$ کدواژهٔ a تعداد یکها در این کدواژه است. ***فاصلهٔ همینگ***[1] $d(a,b)$ بین دو کدواژهٔ a و b تعداد وضعیتهایی است که در آن کدواژه‌ها دارای نمادهای متفاوتند. برای کدواژه‌های

$$a=1010101$$

$$b=1110011$$

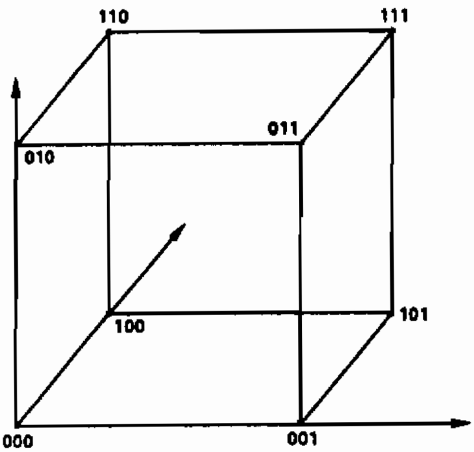
وزنها به ترتیب عبارتند از: $w(a)=4$ و $w(b)=5$ و فاصلهٔ همینگ آنها برابر $d(a,b)=3$ است.

وقتی یک کد را طرح می‌کنیم می‌توانیم کدواژه‌ها را به طریقی انتخاب کنیم که فاصلهٔ همینگ بین هر زوج کدواژه دارای یک مقدار می‌نیمم معیّن باشد. این فاصلهٔ می‌نیمم را ***فاصلهٔ همینگ کد*** می‌نامیم و با d نشان می‌دهیم.

می‌توان کدنمادها را با به‌کارگیری بردارها در یک فضای برداری خطّی نمایش داد. بُعد این فضا برابر تعداد نمادهایی است که کدواژه را می‌سازند. در این صورت برای $n=3$ تصویر نشان داده شده در شکل (۱.۹) برای هشت کدواژهٔ ممکن (۰۰۰) تا (۱۱۱) برای حالت دودویی به دست می‌آید. اکنون اگر کد فقط شامل این هشت کدواژه باشد در این صورت اگر اشتباهی رخ دهد یک کدواژهٔ ارسال شده به یک کدواژهٔ دیگر تبدیل خواهد شد و هیچ تشخیص یا تصحیح خطا امکان‌پذیر نخواهد بود. فاصلهٔ همینگ در این حالت

---
1. Hamming

با ۱ است. اگر کدی با فاصلهٔ همینگ ۲ اختیار کنیم در این صورت کدواژه‌های قابل انتخاب ۴ تاست، یعنی (۰۰۰)، (۰۱۱)، (۱۰۱)، (۱۱۰). اکنون اگر دقیقاً یک خطا رخ دهد آن را می‌توان تشخیص داد چون در این صورت واژهٔ دریافتی متفاوت از چهار کدواژه خواهد بود. با وجود این، هنوز هم دو کدواژهٔ ممکن که فاصلهٔ همینگ ۱ نسبت به واژهٔ دریافت شده دارند وجود دارد. مثلاً، کدواژهٔ دریافت شدهٔ (۱۱۱) می‌توانند از (۰۱۱) یا (۱۰۱) بیاید. بنابراین تصحیح امکان ندارد. تصحیح خطا تنها برای $d=۳$ ممکن است، مثلاً تنها با به‌کارگیری کدواژه‌های (۰۰۰) و (۱۱۱). در این حالت اگر دقیقاً یک خطا رخ دهد دقیقاً یک کدواژه با کمترین فاصلهٔ همینگ وجود دارد و در این صورت، می‌توان نتیجه گرفت که این بایستی کدواژهٔ ارسال شده باشد.



شکل ۱.۹- نمایش سه بعدی کدواژه‌ها به طول سه

از این مثال روشن است که تعداد پیامهای داده شدهٔ $M=۲^k$ به طول n از واژه‌های کانال و همچنین تعداد بررسی $(n-k)$ برابری به اجرای تشخیص خطا یا تصحیح خطای مطلوب کد، و بنابراین به می‌نیمم فاصلهٔ همینگ لازم واژه‌های کانال مربوط می‌شود. به طور کلی، یک کد با فاصلهٔ همینگ d قادر به تشخیص $(d-۱)$ خطا و تصحیح $t=[(d-۱)/۲]$ خطا می‌باشد که در آن $[a]$ بزرگترین عدد صحیح کمتر از a را نشان می‌دهد. بالعکس، کدی که باید t خطا را تصحیح کند بایستی فاصلهٔ همینگ حداقل $۲t+۱$ داشته باشد.

مثال ۱.۹

کد زیر با $n=۵$ و $k=۲$ را در نظر بگیرید:

کدواژه	پیام
(۰۰۰۰۰)	(۰۰)
(۱۰۱۱۰)	(۱۰)
(۰۱۱۰۱)	(۰۱)
(۱۱۰۱۱)	(۱۱)

چون فاصلهٔ همینگ برابر $d = ۳$ است، این کد می‌تواند دو خطا را تشخیص و فقط یک خطا را تصحیح کند. ▲

کد بلوکی (دودویی) خطی C را داریم اگر جمع به پیمانه ۲ از هر زوج کدواژه مجدداً یک کدواژه باشد و اگر هر کدواژه‌ای در ۰ یا ۱ ضرب شود نیز یک کدواژه حاصل شود. از این دو شرط، مستقیماً نتیجه می‌شود که بردار صفر به خودی خود یک کدواژه است. علاوه بر این، نتیجه می‌شود که فاصلهٔ همینگ می‌نیمم یک کد C متناظر با کوچکترین وزن کدواژه‌ها غیر از یک به بردار صفر نسبت داده می‌شود:

$$d = \min\{w(x) \mid x \in C, x \neq 0\}, \tag{۲.۹}$$

که در آن x یک کدواژه را نشان می‌دهد.

این را می‌توان به صورت زیر نشان داد. برای فاصلهٔ همینگ کد داریم

$$d = \min\{d(a,b) \mid a,b \in C, a \neq b\}$$

$$= \min\{d(0,a+b) \mid a,b \in C, a \neq b\}.$$

چون $a+b$ نیز یک کدواژهٔ C است، مثلاً x و چون $d(0,x) = w(x)$ معادلهٔ (۲.۹) مستقیماً به دست می‌آید.

اکنون بررسی می‌کنیم که چگونه می‌توان کدهای بلوکی خطی را طرّاحی نمود. به عنوان مثال یک کد (۷,۴) با طول $n = ۷$ و نمادهای اطلاع (پیام) $k = ۴$ به کار خواهیم برد. از این رو بررسی $n - k = ۳$ برابری وجود دارد.

یک کد (n,k) با کمک k واژه به طور خطی مستقل هر یک به طول n تولید می‌شود. می‌توان این را به کمک ماتریس مولّد G، که در آن سطرها توسط این k واژه تشکیل شده‌اند، نشان داد. اکنون یک کدواژهٔ x بر طبق رابطهٔ زیر به دست می‌آید

$$x = u \cdot G \tag{۳.۹}$$

که در آن u پیامی است که شامل نمادهای اطلاع u_1، u_2، ... ، u_k می‌باشد و G ماتریس مولّد $k \times n$ را نشان می‌دهد. مثالی از ماتریس مولّد G برای کد (۷,۴) ماتریس زیر است:

$$G = \begin{bmatrix} 1000 & 011 \\ 0100 & 101 \\ 0010 & 110 \\ 0001 & 111 \end{bmatrix}.$$

پیام $u = (0111)$ کدواژهٔ زیر را می‌دهد

$$x = ([0\times1+1\times0+1\times0+1\times0],[0\times0+1\times1+1\times0+1\times0],\dots)$$

$$= (0,1,1,1,1,0,0) = (0111100).$$

در این جا جمع با پیمانهٔ ۲ به کار برده شده است $(0+0=0, 0+1=1, 1+0=1, 1+1=0)$. در زمینهٔ کدگذاری معمولاً کدواژه‌ها را با بردار سطری نشان می‌دهند نه به صورت بردار ستونی همچنان که عموماً در پردازش سیگنالها و جبرخطی انجام می‌شود. این مفهوم را در این جا هم دنبال خواهیم کرد.

از این رو کدگذاری پیام را می‌توان به صورت حاصل‌ضرب یک بردار $1\times k$ (پیام) و یک ماتریس مولّد G، $k \times n$ در نظر گرفت. این ماتریس باید شامل ردیفهایی به طور خطّی مستقل باشد و برای کانالهای بی‌حافظهٔ گسسته همواره می‌توان آن را به فرم طبیعی به صورت زیر نوشت

$$G = [I_k , A] \tag{۴.۹}$$

که در آن I_k یک ماتریس همانی $k \times k$ را نشان می‌دهد. این کار قبلاً برای ماتریس مولّد داده شدهٔ قبلی برای کد (۷,۴) انجام شده است. کدی را که در آن ماتریس مولّد دارای چنین فرم طبیعی باشد یک کد سیستماتیک می‌نامیم و دارای این ویژگی است که k نماد اوّل کدواژه با پیام یکی هستند. $n-k$ نماد باقی‌مانده بررسی توازن c_iها هستند و از زیرماتریس A نتیجه می‌شوند

$$[x_1,\dots,x_k] = [u_1,\dots,u_k],$$

$$[x_{k+1},\dots,x_n] = [u_1,\dots,u_k].A. \tag{۵.۹}$$

بنابراین تولید کدواژه‌ها با کد سیستماتیک ساده‌تر است تا با کد غیرسیستماتیک.

با ماتریس مولّد موجود برای بررسی توازن به دست می‌آوریم

$$c_1 = u_2 + u_3 + u_4 ,$$

$$c_2 = u_1 + u_3 + u_4,$$

$$c_3 = u_1 + u_2 + u_4.$$

از این رو برای کدواژهٔ x نتیجه می‌شود که

$$x = (u_1, u_2, u_3, u_4, u_2 + u_3 + u_4, u_1 + u_3 + u_4, u_1 + u_2 + u_4).$$

پیام $u = (0111)$ را با کدواژهٔ $x = (0111100)$ مقایسه کنید.

روشی که در آن بررسی توازن شکل گرفته است در واقع با ماتریسی موسوم به ***ماتریس بررسی توازن*** H تعیین می‌شود. برای C یک کدخطی (n,k) می‌توان یک ماتریس $(n-k)\times n$ یافت که در آن هر سطر یک نماد توازن را مشخص می‌کند. برای کد $(7,4)$ دادهٔ شدهٔ قبلی داریم

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

از این ماتریس بررسی توازن مجدداً نتیجه می‌شود: $c_1 = u_2 + u_3 + u_4$، $c_2 = u_1 + u_3 + u_4$ و $c_3 = u_1 + u_2 + u_4$.

با بازنویسی آنها نتیجه می‌شود

$$u_2 + u_3 + u_4 + c_1 = 0,$$

$$u_1 + u_3 + u_4 + c_2 = 0,$$

$$u_1 + u_2 + u_4 + c_3 = 0.$$

بدیهی است که ماتریس بررسی توازن به قسمی است که حاصل‌ضرب ماتریس کدواژه در یک کدواژهٔ x نتیجه می‌دهد:

$$xH^T = 0. \tag{6.9}$$

به نظر می‌رسد که همهٔ کدواژه‌های تولید شده توسط G در شرط $x.H^T = 0$ صدق می‌کنند. این، رابطهٔ بین ماتریسهای G و H را می‌دهد. دو بردار را متعامد گوییم اگر حاصل‌ضرب اسکالر (داخلی) آنها برابر صفر باشد.

برای هر G ماتریس مولّد $k\times n$ با k سطر مستقل خطّی می‌توان H یک ماتریس $(n-k)\times n$ به دست آورد که در آن $(n-k)$ سطر به طور خطّی مستقلّند و هر کدواژهٔ تولید شده توسط G متعامد نسبت به همهٔ سطرهای H است، یا به عبارت

دیگر $x \cdot h_i = 0$، که h_i یک سطر H است. گاهی اوقات گفته می‌شود که زیرفضای گسترش یافته توسط G بر زیرفضای گسترش یافته توسط H متعامد است. برای یک کد سیستماتیک با ماتریس مولّد $G=[I_k, A]$، ماتریس بررسی توازن به صورت زیر می‌باشد

$$H=[A^T, I_{n-k}]. \quad (۷.۹)$$

با به کار بردن رابطهٔ $x \cdot H^T = 0$ مستقیماً نتیجه می‌شود که مجموع دو کدواژه نیز یک کدواژه است.

$$\begin{aligned}(x+x') \cdot H^T &= x \cdot H^T + x' \cdot H^T \\ &= 0+0=0.\end{aligned}$$

۳.۹ کدگذاری عارضه

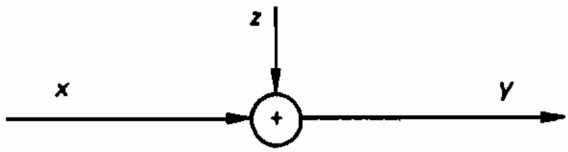
اکنون یک کانال بی‌حافظهٔ گسسته را بررسی می‌کنیم. این فرض که کانال بی‌حافظه است به این معناست که فرض می‌کنیم خطا در لحظات دلخواهی بدون تأثیر بر یکدیگر می‌تواند رخ دهد. اگر یک واژهٔ y دریافت شده باشد آن را به عنوان مجموع کدواژهٔ ارسال شدهٔ x و یک بردار خطای z در نظر می‌گیریم. از این رو یک کانال جمعی را بررسی می‌کنیم (شکل (۲.۹) را ببینید).

بردار خطا در محلهایی که در آن x و y متفاوتند شامل ۱ می‌باشد، از این رو خطایی در زمان ارسال رخ داده است. برای کدگشایی از ماتریس بررسی توازن H استفاده می‌کنیم، چون برای کدواژهٔ x، $xH^T = 0$. اگر خطا رخ داده باشد در این صورت $z \neq 0$، لذا منطقی است که $yH^T \neq 0$. بردار $S = y \cdot H^T$ را ***عارضه*** می‌نامیم. عارضه به x بستگی ندارد بلکه تنها به z بستگی دارد چون:

$$\begin{aligned}S &= y \cdot H^T \\ &= (x+z) \cdot H^T \\ &= x \cdot H^T + z \cdot H^T \\ &= z \cdot H^T. \end{aligned} \quad (۸.۹)$$

بنابراین اگر $S=0$ به سادگی می‌توان با بازبینی کردن خطا را مشخّص کرد. اگر $z=0$ این نیز حالتی است که $S=0$. نمی‌توان از $S=0$ با اطمینان نتیجه گرفت که $z=0$؛ بعضی بردارهای خطا اجازه می‌دهند که کدواژه‌ای به کدواژهٔ دیگر منتقل شود

که به عنوان نتیجه $S=0$ را می‌دهد و از این رو یک خطای غیر قابل تشخیص را می‌دهد.



شکل ۲.۹ کانال جمعی

اگر خطایی تشخیص داده شد، بعداً باید مشخّص شود که در کجا رخ داده است. بنابراین بایستی الگوی خطا را از S بازسازی کنیم. در این جا خود را به کدهایی که قادر به تصحیح یک خطا در هر کدواژه‌اند محدود می‌کنیم، برای مثال داده شدهٔ یک کد (۷,۴) در این صورت هشت نمونه خطای z ممکن وجود دارد؛ یعنی هیچ خطا ($z=0$) و هفت مرتبه یک خطا (۰۰۰۰۰۰۱) تا (۱۰۰۰۰۰۰).

عارضهٔ $S=y.H^T$ یک بردار (۱×۳) با هشت مقدار ممکن است: (۰۰۰) تا (۱۱۱). در این حالت می‌بینیم یک رابطهٔ یکتایی بین یک بردار خطا و مقدار عارضهٔ ممکن است. اکنون می‌توان پس از تعیین S با یافتن بردار خطای متناظر در جدول و اضافه کردن آن به واژهٔ دریافت شده y (به پیمانهٔ ۲) تصحیح را انجام داد. رابطهٔ بین z و S با افزودن دائمی یکی از نمونه‌های خطا، z، به کدواژهٔ (۰۰۰۰۰۰۰) و محاسبهٔ این که S چیست نتیجه می‌شود. در این روش جدول داده شدهٔ زیر را به دست می‌آوریم

S	z
۰۰۰	۰۰۰ ۰۰۰ ۰
۰۰۱	۰۰۰ ۰۰۰ ۱
۰۱۰	۰۰۰ ۰۰۱ ۰
۱۰۰	۰۰۰ ۰۱۰ ۰
۱۱۱	۰۰۰ ۱۰۰ ۰
۱۱۰	۰۰۱ ۰۰۰ ۰
۱۰۱	۰۱۰ ۰۰۰ ۰
۰۱۱	۱۰۰ ۰۰۰ ۰

مثال ۹. ۲

فرض کنید $y = (0010010)$ دریافت شده است. در این صورت عارضه عبارت است از

$$S = (0010010)\begin{bmatrix} 011 \\ 101 \\ 110 \\ 111 \\ 100 \\ 010 \\ 001 \end{bmatrix} = (100).$$

از این رو، نتیجه می‌شود که y یک واژه نیست. از جدول داده شده در بالا نتیجه می‌شود که $S = (100)$ متناظر با بردار خطای $z = (0000100)$ می‌باشد. با جمع کردن y و z کدواژه کدگشایی شده صحیح نتیجه می‌شود

$$x = (0010110).$$

▲

در مثال داده شده در این جا تعداد مقادیر عارضه برابر با تعداد بردارهای خطا با وزن حداکثر یک می‌باشد.

یک روش کلّی‌تر از مفهوم هم‌مجموعه استفاده می‌کند. برای C یک کد (n,k) و یک واژهٔ دلخواه a به طول n مجموعهٔ $a+C$ به صورت زیر

$$a + C = \{a + x \mid x \in C\} \tag{9.9}$$

را یک هم‌مجموعهٔ C می‌نامند. همهٔ اعضای هم‌مجموعه دارای عارضهٔ یکسانی هستند چون

$$(a+x).H^T = a.H^T + x.H^T = a.H^T.$$

برای عارضهٔ داده شدهٔ $S = z.H^T$ جوابهای ممکن z به هم‌مجموعهٔ C منجر می‌شود. 2^{n-k} تا از این هم‌مجموعه‌ها هر یک متناظر با یکی از 2^{n-k} عارضهٔ ممکن وجود دارد. هر هم‌مجموعه شامل 2^k عضو است. از این رو در مجموع $2^{n-k}.2^k = 2^n$ عضو که متناظر با 2^n واژه دریافت شده ممکن y است وجود دارد. از این رو وقتی گیرنده عارضه را تعیین کرد، جست‌وجو برای بردار خطای صحیح (و از این رو کدواژهٔ ارسال شده) واقعاً از 2^n به 2^k امکان برگردانده می‌شود.

برای تعیین بردار خطای مناسب از این 2^k امکان محتمل‌ترین بردار را تعیین خواهیم کرد. برای کانال بی‌حافظه متقارن دودویی با احتمال خطای p در نظر گرفته شده در این جا، احتمال یک بردار خطای معیّن برابر است با

$$p(z)=\prod_{i=1}^{n} p(z_i),$$

که در آن

$$p(z_i=0)=1-p,$$

$$p(z_i=1)=p.$$

اگر بردار خطا شامل t خطا باشد در این صورت احتمال چنین بردار خطایی برابر است با

$$p(z)=p^t(1-p)^{n-t}.$$

چون معمولاً $p<\frac{1}{2}$، $p(z)$ تابع نزولی از t است بنابراین محتملترین بردار خطای z برداری با کمترین مقدار t است، یعنی با کمترین وزن می‌باشد. از این رو به الگوریتم کدگشایی زیر می‌رسیم:

- عارضهٔ $S=y\,.\,H^T$ را محاسبه می‌کنیم.
- بردار می‌نیمم وزن z را در هر مجموعهٔ متناظر با S به دست می‌آوریم.
- فرض می‌کنیم $\hat{x}=y-z$.

روشن خواهد شد که مرحلهٔ دوم پیچیده‌ترین است. اگر k و $(n-k)$ نسبتاً کوچک باشند می‌توان روشی موسوم به روش جست‌وجوی جدولی را دنبال کرد. این روش را با کمک کد (۵,۳) بررسی خواهیم کرد. ماتریس مولّد این کد عبارت است از

$$G=\begin{bmatrix}1&0&0&1&1\\0&1&0&0&1\\0&0&1&0&1\end{bmatrix},$$

در حالی که ماتریس بررسی توازن به صورت زیر می‌باشد

$$H=\begin{bmatrix}1&0&0&1&0\\1&1&1&0&1\end{bmatrix}.$$

چهار عارضه یعنی ۰۰، ۰۱، ۱۰، ۱۱ وجود دارد. گیرنده می‌تواند $32=2^5$ واژه دریافت کند که آن را در یک ماتریس 8×4 منظّم می‌کنیم. هر ردیف شامل واژه‌هایی است که در $S=z\,.\,H^T$ برای عارضهٔ معیّنی صدق می‌کند. ماتریس با جدول زیر داده شده است و آن را ***ماتریس استاندارد*** می‌نامیم

S	هم‌مجموعهٔ راهنما							
۰۰	۰۰۰۰۰	۰۰۱۰۱	۰۱۰۰۱	۰۱۱۰۰	۱۰۰۱۱	۱۰۱۱۰	۱۱۰۱۰	۱۱۱۱۱
۰۱	۰۰۰۰۱	۰۰۱۰۰	۰۱۰۰۰	۰۱۱۰۱	۱۰۰۱۰	۱۰۱۱۱	۱۱۰۱۱	۱۱۱۱۰
۱۰	۰۰۰۱۰	۰۰۱۱۱	۰۱۰۱۱	۰۱۱۱۰	۱۰۰۰۱	۱۰۱۰۰	۱۱۰۰۰	۱۱۱۰۱
۱۱	۱۰۰۰۰	۱۰۱۰۱	۱۱۰۰۱	۱۱۱۰۰	۰۰۰۱۱	۰۰۱۱۰	۰۱۰۱۰	۰۱۱۱۱

سطرهای این ماتریس هم‌مجموعهٔ C هستند. سطر اوّل خود کد است. بردار با کمترین وزن در داخل هم‌مجموعه را در سمت چپ قرار داده و آن را هم‌مجموعهٔ راهنمای $z(s)$ می‌نامیم. به استثنای سطر اوّل، یک بردار در هم‌مجموعه برابر با هم‌مجموعهٔ راهنما به اضافهٔ کدواژهٔ بالای بردار سؤال می‌باشد. بنابراین ۰۱۱۰۱ در سطر دوم برابر مجموع ۰۰۰۰۱ و ۰۱۱۰۰ می‌باشد.

اکنون مرحلهٔ دوم الگوریتم کدگشایی ساده می‌شود. عارضه هم‌مجموعه را تعیین می‌کند. چون هم‌مجموعهٔ راهنما $z(s)$ هم‌مجموعه کمترین وزن را دارد و چون محتمل‌ترین است به عنوان الگوی خطایی که برای آن جست‌وجو شده در نظر گرفته می‌شود. در این صورت پیام فرض شده $\hat{x} = y - z(s)$ می‌شود. در حالت کدواژه دریافت شده $y = (۰۱۱۰۱)$ به $\hat{x} = (۰۱۱۰۰)$ منجر می‌شود.

واضح است که تصحیح خطا تنها اگر الگوی خطا واقعاً هم‌مجموعه راهنما باشد درست است.

در حالتهای دیگر که به علّت وزنشان کمتر محتملند تصحیح نادرستی رخ می‌دهد. اگر بخواهیم قادر به تصحیح نمونه‌های خطای بیشتری باشیم، بایستی تعداد مقادیر عارضه بزرگتر شود، که تنها با افزایش تعداد $n-k$ بررسی توازن ممکن می‌شود.

روش کدگشایی برای مقادیر بزرگ n و $n-k$ خیلی پیچیده می‌شود. روشهای دیگری وجود دارند که روشهای کدگشایی نسبتاً ساده‌ای را برای طولهای بزرگ میسّر می‌سازند. یک دستهٔ مهم کدها با کمک کثیرالجمله‌ها از مشخّصات استفاده می‌کنند. این موضوع خارج از بحث این کتاب است و در این جا بیشتر توضیح داده نخواهد شد.

۴.۹ کدهای همینگ

کدهای همینگ دسته‌ای از کدها را تشکیل می‌دهند که می‌توانند یک خطا را تصحیح کنند. ساده‌ترین روش توصیف کردن این کدها با کمک ماتریس بررسی توازن است. یک کد همینگ شامل ماتریس بررسی توازن $r \times (۲^r - ۱)$ است و همهٔ بردارهای $(r \times ۱)$ غیرصفر را به عنوان ستونها دارا می‌باشد. به عنوان نتیجه‌ای از این یک کد همینگ یک کد $(۲^r - ۱, ۲^r - r - ۱)$ با طول $n = ۲^r - ۱$، $k = ۲^r - r - ۱$ و r بررسی توازن می‌باشد.

از این رو برای $r = ۲$ به دست می‌آوریم

$$H_۲ = \begin{bmatrix} ۱ & ۱ & ۰ \\ ۱ & ۰ & ۱ \end{bmatrix}.$$

ماتریس مولّد برای این کد عبارت است از:

$$G_۲ = [۱ \quad ۱ \quad ۱],$$

که همچنین نشان می‌دهد که با یک کد تکراری نیز سروکار داریم. بنابراین، مستقیماً می‌توان دید که یک کد می‌تواند یک خطا را تصحیح کند.

برای $r = ۳$، به همین طریق به دست می‌آوریم

$$H_۳ = \begin{bmatrix} ۰ & ۱ & ۱ & ۱ & ۱ & ۰ & ۰ \\ ۱ & ۰ & ۱ & ۱ & ۰ & ۱ & ۰ \\ ۱ & ۱ & ۰ & ۱ & ۰ & ۰ & ۱ \end{bmatrix},$$

که در آن ستونها طوری تنظیم شده‌اند که یک کد سیستماتیک به دست آمده است. ماتریس مولّد متناظر آن عبارت است از

$$G_۳ = \begin{bmatrix} ۱ & ۰ & ۰ & ۰ & ۰ & ۱ & ۱ \\ ۰ & ۱ & ۰ & ۰ & ۱ & ۰ & ۱ \\ ۰ & ۰ & ۱ & ۰ & ۱ & ۱ & ۰ \\ ۰ & ۰ & ۰ & ۱ & ۱ & ۱ & ۱ \end{bmatrix}.$$

این کد-همینگ (۷,۴) قبلاً در این فصل به عنوان یک مثال به کار برده شده است و نیز می‌تواند یک خطا را تصحیح کند. به خصوص کدگشایی کدهای همینگ یک مزیّت بزرگ این کدهاست، زیرا نسبتاً ساده‌اند. هم‌مجموعهٔ راهنماهای یک کد همینگ دقیقاً $n + ۱ = ۲^r$ بردار با وزن ≤ ۱ می‌باشد به قسمی که همهٔ نمونه‌های خطا با صفر یا یک خطا را می‌تواند تصحیح کند.

با مرتّب کردن ستونهای H ماتریس بررسی توازن به ترتیب صعودی مقدار دودویی، الگوریتم کدگشایی سادهٔ زیر به دست می‌آید

۱. عارضهٔ $S = y.H^T$ را تعیین می‌کنیم.
۲(الف). اگر $S = 0$، آن گاه $\hat{x} = y$.
۲(ب). اگر $S \neq 0$، آن گاه S نمایش دودویی مکان خطا را می‌دهد. (تحت شرطی که دقیقاً یک اشتباه انجام شده باشد).
۳. $\hat{x} = y + z$.

مثال ۳.۹

ماتریس H_3 به صورت زیر مرتّب شده است

$$H_3 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

فرض کنید که $y = [1111011]$ دریافت شده باشد. اکنون عارضه، برابر $S = [101]$ می‌شود و بنابراین برابر با ستون پنجم است. از این رو یک اشتباه در مکان پنجم رخ داده است. اکنون تصحیح به واژهٔ $\hat{x} = [1111111]$ منجر می‌شود.

۵.۹ تمرینها

۱.۹ در کدگذاری یک کانال دودویی کدواژه‌ها از دو نماد اطلاع b_1 و b_2 و سه نماد بررسی توازن (c_1، c_2 و c_3) تشکیل شده است. ماتریس مولّد به صورت زیر می‌باشد

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

(الف) کدواژه‌ها، وزن آنها و فاصلهٔ همینگ آنها را تعیین کنید.
(ب) H ماتریس بررسی توازن را تعیین کنید.
(پ) عارضه را از نمونه‌های خطای زیر به دست آورید:

$$(01000) \quad , \quad (00101) \quad , \quad (10010) \quad , \quad (11111) \; ,$$

(ت) کدواژهٔ (۱۱۰۱۰) تولید شده است، با نمونهٔ خطای زیر دگرشکلی کدام است: (۱۰۰۱۰). کدواژهٔ دریافت شده را تعیین کنید و چه تصحیحی به کار خواهید برد. تصمیم را

شرح دهید.

۲.۹ یک کد دودویی با اضافه کردن سه نماد بررسی توازن به نمادهای اطلاع (نشان داده شده با b_1، b_2 و b_3) ساخته شده است. برای این کد نمادها داریم:

$$x_1 = b_1,$$
$$x_2 = b_2,$$
$$x_3 = b_3,$$
$$x_4 = b_2 + b_3,$$
$$x_5 = b_1 + b_3,$$
$$x_6 = b_1 + b_2,$$

(الف) کدواژه‌های این کد را تعیین کنید.
(ب) H ماتریس بررسی توازن یا کنترل را تعیین کنید.
(پ) G ماتریس مولّد را تعیین کنید.
(ت) فاصلهٔ همینگ کد را به دست آورید.
(ث) این کد چند خطا را می‌تواند تشخیص دهد و چند خطا را می‌تواند تصحیح کند.

۳.۹ یک کد بلوکی خطّی که در آن برای هر سه نماد اطلاع یک نماد بررسی توازن اضافه شده، به قسمی که تعداد «۱» در هر کدواژه زوج است در نظر می‌گیریم.

(الف) H ماتریس کنترل کد را بدهید.
(ب) G ماتریس مولّد کد را به دست آورید.
(پ) فاصلهٔ همینگ کد را تعیین کنید.
(ت) چند خطا می‌تواند تشخیص دهد و چند خطا می‌تواند تصحیح کند؟ جوابتان را شرح دهید.
(ث) ماتریس استاندارد را در رابطه با هم‌مجموعه که نقشی در کدگشایی با کمک عارضه بازی می‌کند به دست آورید.

۴.۹ یک کد بلوکی خطّی دارای ماتریس بررسی توازن زیر است

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

(الف) این کد شامل چند کدواژه است؟
(ب) ماتریس مولّد سیستماتیک کد را به دست آورید.
(پ) پس از ارسال دو کدواژه با کانال متقارن سیستماتیک (با احتمال خطای ۰٫۵ >) بردارهای (۱۰۰۰۰۰۱) = y_1 و (۱۰۰۱۱۰۰) = y_2 دریافت شده‌اند.

- عارضه برای هر دو بردار را محاسبه کنید.
- برای هر دو بردار با کمک عارضهٔ کدگشایی، محتملترین کدواژهٔ ارسال شده را تعیین کنید.
- محتملترین نمادهای اطلاع چیست؟

۶.۹ جوابها

۱.۹ (الف) با به کار بردن $x = u.G$، که u پیامی شامل b_1 و b_2 است، به کدواژه‌های زیر منجر می‌شود

x	$w(x)$
۱۰۱۱۱	۴
۰۱۱۰۱	۳
۱۱۰۱۰	۳
۰۰۰۰۰	۰

فاصلهٔ همینگ بین کدواژه‌ها در ماتریس متقارن زیر نشان داده شده است

$$\begin{bmatrix} ۰ & ۳ & ۳ & ۴ \\ ۳ & ۰ & ۴ & ۳ \\ ۳ & ۴ & ۰ & ۳ \\ ۴ & ۳ & ۳ & ۰ \end{bmatrix}.$$

(ب) ماتریس بررسی توازن را برای هر x می‌توان با $x.H^T = ۰$ به دست آورد. فرم کلّی H عبارت است از

$$H = \begin{bmatrix} a & b & ۱ & ۰ & ۰ \\ c & d & ۰ & ۱ & ۰ \\ e & f & ۰ & ۰ & ۱ \end{bmatrix}.$$

با کمک چهار کدواژه و $x.H^T=0$ به دست می‌آوریم

کدواژهٔ ۱: $a=1$

$c=1$

$e=1$

کدواژهٔ ۲: $b=1$

$d=0$

$f=1$

کدواژهٔ ۳: $a+b=1$

$c+d=1$

$e+f=0$

کدواژهٔ ۴: ——

ماتریس H حاصل برابر است با

$$H=\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

(پ) اگر z یک بردار خطا باشد، عارضهٔ S عبارت است از $S=z.H^T$. برای چهار بردار خطای ذکر شده به دست می‌آید $S=[101]$.

(ت) کدواژهٔ ارسال شدهٔ x عبارت است از (۱۱۰۱۰)، بردار خطا برابر است با (۱۰۰۱۰)، از این رو کدواژهٔ دریافت شده $y=x+z=(01000)$ با عارضهٔ داده شده در c می‌باشد.

برای بردارهای داده شده در c، (۰۱۰۰۰) دارای کمترین وزن است. در نتیجه این بردار را به عنوان بردار خطا در نظر می‌گیریم. بنابراین، کدواژهٔ مفروض $\hat{x}=y-z(s)=(00000)$ می‌شود که در این حالت نادرست است.

۲.۹ (الف) ۸ پیام وجود دارد. کدواژه‌های متناظر آنها به صورت زیر داده شده‌اند

کدواژه	پیام
۰۰۰ ۰۰۰	۰۰۰
۰۰۱ ۱۱۰	۰۰۱
۰۱۰ ۱۰۱	۰۱۰
۰۱۱ ۰۱۱	۰۱۱
۱۰۰ ۰۱۱	۱۰۰
۱۰۱ ۱۰۱	۱۰۱
۱۱۰ ۱۱۰	۱۱۰
۱۱۱ ۰۰۰	۱۱۱

(ب) ماتریس کنترل H برای هر کدواژهٔ x باید در $x.H^T=0$ صدق کند، نتیجه می‌شود که

$$H=\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(پ) ماتریس مولّد G را از $G.H^T=0$ می‌توان به دست آورد. نتیجه عبارت است از

$$G=\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

این نتیجه نیز می‌تواند با نوشتن H به صورت $H=[A^T, I_{n-k}]=[A^T, I_3]$ به دست آید و در این صورت توجّه کنید که $G=[I_3, A]$. امکان دیگر این است که G را از $x=u.G$ نتیجه بگیریم.

(ت) کوچکترین فاصلهٔ همینگ بین کدواژه‌ها ۳ است. از این رو فاصلهٔ همینگ d کد برابر ۳ است.

(ث) برای فاصلهٔ همینگ $d=3$ مقدار $d-1=2$ خطا می‌توان تشخیص داد و $t=[(d-1)/2]=1$ خطا می‌توان تصحیح کرد.

۳.۹ (الف) واضح است که برای نماد بررسی توازن $c_1=u_1+u_2+u_3$ برقرار است. از این رو ماتریس کنترل به صورت زیر داده شده است

$$H = [1111]$$

(ب) اگر $H = [A^T, I_{n-K}]$ در این صورت $G = [I_k, A]$. در این جا، ماتریس مولّد به صورت زیر می‌باشد

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

(پ) کدواژه‌ها به صورت جدول زیر داده شده‌اند

کدواژه x	پیام u
۰۰۰۰	۰۰۰
۰۰۱۱	۰۰۱
۰۱۰۱	۰۱۰
۰۱۱۰	۰۱۱
۱۰۰۱	۱۰۰
۱۰۱۰	۱۰۱
۱۱۰۰	۱۱۰
۱۱۱۱	۱۱۱

برای این فاصلهٔ همینگ $d = 2$ نتیجه می‌شود.

(ت) چون $d = 2$، ۱ $(= d - 1)$ خطا می‌توان تشخیص داد. هیچ خطایی را نمی‌توان تصحیح کرد چون $t = [(d-1)/2] = 0$.

(ث) در این حالت داریم $S = y.H^T$. دو عارضهٔ $S = 0$ و $S = 1$ وجود دارند. به دست می‌آوریم

S	هم‌مجموعهٔ راهنما							
۰	۰۰۰۰	۰۰۱۱	۰۱۱۰	۱۱۰۰	۰۱۰۱	۱۰۱۰	۱۰۰۱	۱۱۱۱
۱	۰۰۰۱	۰۰۱۰	۰۱۱۱	۱۱۰۱	۰۱۰۰	۱۰۱۱	۱۰۰۰	۱۱۱۰

۴.۹ (الف) در این جا ممکن است یادآوری کرد که با در نظر گرفتن ماتریس H ممکن است نتیجه گرفت که با کد همینگ مواجهیم. ماتریس بررسی توان به صورت یک ماتریس $(n-k)\times n$ است. از این رو $n=۷$ و $n-k=۷-k=۳$ که نتیجه می‌شود $k=۴$. اکنون $۲^k=۱۶$ پیام وجود دارد.

(ب) ماتریس مولّد G به دست آمده از H عبارت است از

$$G=\begin{bmatrix} ۱ & ۰ & ۰ & ۰ & ۱ & ۱ & ۱ \\ ۰ & ۱ & ۰ & ۰ & ۱ & ۱ & ۰ \\ ۰ & ۰ & ۱ & ۰ & ۱ & ۰ & ۱ \\ ۰ & ۰ & ۰ & ۱ & ۰ & ۱ & ۱ \end{bmatrix}$$

(پ) برای کدواژهٔ $y_۱$ عارضه عبارت است از $S_۱=y_۱.H^T=(۱۱۰)$. این متناظر با دومین ستون ماتریس بررسی توازن H می‌باشد. از این رو، احتمالاً بیت دوم کدواژهٔ دریافت‌شده باید تصحیح گردد. بنابراین کدواژهٔ مفروض (۱۱۰۰۰۰۱) و پیام ارسال شده (۱۱۰۰) می‌باشد.

کدواژهٔ $y_۲$ دارای عارضهٔ $S_۲=y_۲.H^T=(۰۰۰)$ است. به موجب این مقدار عارضه، فرض شده است که خطایی وجود ندارد و کدواژهٔ دریافت‌شده درست است. پیام ارسال شده (۱۰۰۱) می‌باشد.

۱۰

رمزشناسی

۱.۱۰ رمزنگاری و تحلیل رمز

زمینهٔ علمی دیگری که در آن نتایج نظریهٔ اطلاع به کار برده شده، ***رمزشناسی*** است چون کلمهٔ رمزشناسی اختصاری از واژه‌های یونانی «کرپتو[1]» و «لوگو[2]» است، دقیقاً به معنی مطالعهٔ پنهان‌کاریهاست. همچنان که بین دیگر روشها به توسعهٔ روشهایی برای ***رمزگذاری*** و ***رمزگشایی*** پیامها می‌پردازد.

خواستن این که پیامها را به فرمی درآوریم که افرادی که پیام به آنها ربطی ندارد نتوانند پیام اصلی را بفهمند بسیار قدیمی است. از زمان شروع تاریخ ثبت شده از کدهای سرّی استفاده می‌شده است. از تاریخ اوّلیه تا حدود جنگ جهانی دوم کدهای سرّی اصولاً در ارتش و محیطهای سیاسی به کار برده می‌شد. در سالهای اخیر پیشرفتهایی رخ داده است که موجب شده تقاضا برای روشهای رمزی‌کردن پیامها خارج از ارتش یا محیط سیاسی نیز گسترش یابد.

برای ارائهٔ یک مثال، کابل تلویزیون را که در اصل تنها مشتری مشترک باید قادر به دریافت برنامه‌های ارسال شده توسط ایستگاه تلویزیون باشد در نظر بگیرید، در حالی که دریافت برای آنهایی که مشترک نیستند باید غیرممکن باشد. به این کار می‌توان نایل شد اگر تصاویر ارسال شده به صورت رمزی شده باشند در حالی که سیستمی در گیرنده قرار داده شود که تصویر را مجدداً رمزگشایی کند. حوزهٔ مهمی که نمی‌تواند بدون رمزشناسی وجود داشته باشد عبارت از اجزای موقّتی بانک الکترونیکی، کارتهای مغناطیسی و

1. Kruptos 2. Logos

قبولیهای بانکی و غیره است که همهٔ آنها ابزارهای رمزنگاری را به کار می‌برند.

به مسائلی که مربوط به محافظت شخصی است فکر کنید. به علّت گسترش حالت خودکار، تعداد دائماً در حال افزایش سیستمهایی که شامل بانک داده‌ها با اطلاعات شخصی است وجود دارد، مثلاً، بانک داده‌های پزشکی (طبی)، بانک داده‌های قضایی و غیره در بسیاری از این حالتها مطلوب آن است که داده‌های ذخیره شده در برابر مشاوره ناخواسته محافظت شوند. ممکن است همچنین بخواهیم در زمان انتقال داده‌ها (متن، گفتار، ویدیو) از طریق شبکه‌های ارتباطی، مانع استراق سمع شویم.

در رمزشناسی بین دو شیوه یعنی ***رمزنگاری*** و ***تحلیل رمزی*** وجه تمایزی در نظر گرفته می‌شود. رمزنگاری بخشی از رمزشناسی است که مربوط به توسعه و مطالعهٔ روشها و روشهای رمزگذاری می‌شود. در این جا معمولاً می‌توان از کلیدهای سرّی استفاده کرد. تنها کسانی که کلید سرّی را دارند می‌توانند اطلاعات رمزی شده را رمزگشایی کنند؛ برای افراد دیگر، این کار بسیار دشوار و تقریباً غیرممکن است. تحلیل رمزی بخشی از رمزشناسی برای توسعهٔ تکنیکها رمزگشایی پیامهای رمزشده است؛ یعنی این که بدون هیچ آگاهی قبلی نسبت به کلید «خاصّی» جست‌وجو می‌کند.

بدیهی است که پرداختن به صور مختلف رمزنگاری و تجزیه و تحلیل رمزی از حوصلهٔ این کتاب خارج است؛ بنابراین در این فصل به طور اجمالی به این موضوع می‌پردازیم و به آن جنبه‌هایی از رمزنگاری و تجزیه و تحلیل رمزی توجّه می‌کنیم که در آنها به کاربرد مفاهیم یافته شده درنظریهٔ اطلاع تأکید می‌شود.

۲.۱۰ طرح کلّی سیستمهای رمزی

در شکل (۱.۱۰) خلاصهٔ کلّی از یک ***سیستم رمزی*** داده شده است. منبعی که پیامهای M را تولید می‌کند در کنار فرستنده وجود دارد که با ***متن ساده*** نشان داده شده است. متن ساده با استفاده از برخی از روشهای رمزگذاری به ***متن رمزی*** انتقال داده می‌شود که با حرف C نشان داده شده است. عمل رمزی کردن را می‌توان به صورت تبدیل T در نظر گرفت که M را به C تبدیل می‌کند. تعدادی مثال ساده از روشهای رمزی کردن را در بخشهای بعدی ارائه خواهیم کرد.

برخی امکانات برای تبدیل وابسته به انتخاب کلید K وجود دارد. این کلید از مجموعه کلیدهای ممکن تولید می‌شود بنابراین داریم

$$C = T_K(M).$$

رمزگشایی در طرف دریافتی با پردازش متن رمزی با تبدیل وارون T_K^{-1} رخ می‌دهد، یعنی داریم $M = T_K^{-1}(C)$. معمولاً فرض می‌شود که تبدیل خودش معلوم است، ولی کلید معلوم نیست.

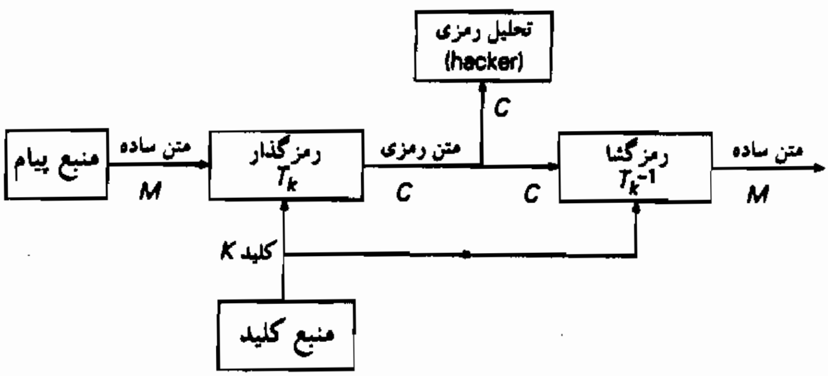
هدف تحلیلگر کشف کلید متن رمزی یا کشف مستقیم متن ساده است. در آینده فرض می‌کنیم که کلید برای هر پیام جدید متفاوت است. یعنی متن ساده، متن رمزی و کلید همه را می‌توان به صورت کمیّت تصادفی در نظر گرفت. در این جاست که می‌توانیم پیوندی با نظریهٔ اطلاع، همان‌طور که می‌خواهیم، ایجاد کنیم.

سه نوع اقدام تحلیلگرانه روی سیستم رمزی متناظر با طبیعت اطلاع که تحلیلگر باید به آن دسترسی پیدا کند می‌توان صورت داد. این سه نوع اقدام عبارتند از

- اقدام فقط با متن رمزی
- اقدام با متن ساده-معلوم
- اقدام با متن ساده-انتخابی

در رویدادی که تحلیلگر فقط متن رمزی را دارد (پیام رمزی شده)، باید سعی کند پیام واقعی (متن ساده) را با تجزیه و تحلیل کردن ساختمان و جنبه‌های آماری ممکن موجود در متن رمزی، رمزگشایی کند یا مهمتر این که باید سعی کند کلید را بیابد. این کار اقدام فقط با متن رمزی است.

در وضعیتی که در آن علاوه بر داشتن متن رمزی تحلیلگر دربارهٔ متن سادهٔ نظیر نیز اطلاعی دارد خیلی رضایت‌بخشتر از وضعیت قبلی است. تحلیلگر اکنون از طریق آگاهی از



شکل ۱.۱۰- خلاصهٔ یک سیستم رمزی

ترکیباتی از قطعاتی از متن رمزی و متن ساده می‌تواند سعی کند قسمتی از متن رمزی را که برای آن متن ساده نظیر معلوم نیست کدگشایی کند (اقدام با متن ساده-معلوم). وقتی تحلیلگر موفق به نفوذ در سیستم رمزی یا کاربر سیستم شده باشد وضعیتهایی که در آن تحلیلگر متن رمزی و همچنین بخشی از متن ساده متناظر را دارد اتفاق می‌افتد. مثلاً برای دادوستد موقتی خودکار اطلاع راجع به شخصی که پول را انتقال می‌دهد و بانکی که باید پول به آن منتقل شود در هر معامله باید موجود باشد. اگر به واسطهٔ اطلاعات جنبی تحلیلگر بداند در کجا اطلاعات راجع به بانک، شماره حساب و غیره در متن رمزی مخفی شده‌اند، در این صورت می‌تواند بکوشد بقیهٔ متن رمزی را براساس این آگاهی رمزگشایی کند.

رضایتبخشترین وضعیت برای تحلیلگر وقتی اتفاق می‌افتد که تحلیلگر می‌تواند متن ساده را خودش انتخاب کند و می‌تواند این متن را با متن رمزی حاصل مقایسه کند (اقدام با متن ساده-انتخابی). این وضعیت می‌تواند در حالت تحلیل سیستم رمزی برای پردازش واژه در بین وضعیتهای دیگر رخ دهد.

وقتی روشهای رمزنگاری به کار برده می‌شوند طبیعتاً اصلاح سیستم ترجیح داده می‌شود که دلیلی در مقابل هر سه نوع اقدام است. در عمل تشخیص این دلایل مشکل می‌باشد. سیستمی که به نظر می‌رسد در مقابل اقدام فقط با متن رمزی در امان است لزومی ندارد که در مقابل اقدامی به روش متن ساده چنین باشد. در عمل، سیستمی که می‌تواند اقدامی را براساس متن ساده-انتخابی باقی نگه‌دارد در سطح بالاتری از سیستمی که می‌تواند اقدامی را فقط براساس متن رمزی باقی نگه‌دارد مورد توجه قرار می‌گیرد.

تعدادی از کاربردهایی که در آن رمزنگاری به کار رفته است در بخش قبل آورده شد. به طور کلّی، کاربردها را می‌توان به دو گروه یعنی کاربرد مربوط به ذخیره‌سازی و کاربرد راجع به ارسال تقسیم کرد.

در حالت ذخیره‌سازی باید به ذخیرهٔ داده‌ها در سیستم رایانه‌ای: روی دیسکت یا روی نوار مغناطیسی اندیشید. در این حالت اغلب روشی که از طریق آن داده‌ها یا نرم‌افزار ذخیره می‌شود، شناخته شده‌اند، ولی کلید شناخته شده نیست. در این جا اقدام تحلیلگر جالب است زیرا اغلب داده‌ها برای مدّت زمان طولانی ذخیره شده‌اند. از این رو تحلیلگر برای یافتن کلید نمونه زمانی دارد.

در حالت ارسال (تلفن، تلویزیون، از طریق کابل یا ارتباط ماهواره‌ای) معمولاً تنها پیام رمزی شده برای مدّت خیلی کوتاهی از زمان برای تحلیلگر موجود است. علاوه بر

این، تعویض کلید خیلی ساده‌تر از حالت ذخیره رخ می‌دهد. به علاوه، در ارتباط کابلی، پیامها اغلب تنها برای مدّت زمان محدودی با ارزشند، زیرا محتوا ممکن است پس از مدّت زمان معیّنی تاریخ گذشته باشد، برای مثال، به اخبار، اطلاعات هوا و غیره فکر کنید). روش رمزنگاری که ممکن است مطلقاً امن نباشد هنوز هم ممکن است جالب به نظر برسد به شرط آن که مقدار پیام در محدودهٔ مقدار زمانی که یک استراق سمع کننده برای رمزگشایی لازم دارد خیلی کاهش یافته باشد.

۳.۱۰ سیستم رمزی

دو روش اساسی رمزی کردن می‌توان تشخیص داد

- ***رمزی کردن گروهی***
- ***رمزی کردن بلوکی***

در رمزی کردن گروهی، پیام به صورت دنباله‌ای متوالی که از تعدادی عضو جدا ساخته شده است در نظر گرفته می‌شود. برای مثال می‌توان فکر کرد که اعضا از حروف و هم دودویی یا به صورت ASCII باشند. مشخصهٔ رمزی کردن گروهی آن است که متن را عضو به عضو رمزی می کنند. با این روشِ رمزی کردن اغلب ثبت نوبتی به کار برده می‌شود.

در حالت رمزی کردن بلوکی تعدادی عضو به طور پیوسته با یکدیگر انتخاب می‌شوند و به عنوان یک واحد رمزی می‌شوند. یک مثال از رمزی کردن بلوکی الگوریتم *DES* توسعه یافته توسط *IBM* در سالهای ۱۹۶۸ تا ۱۹۷۵ می‌باشد (داده‌های رمزی شده استاندارد = *DES*)، که در حال حاضر یکی از الگوریتمهایی است که برای رمزی کردن بلوکی در سطح بسیار وسیعی به کار برده می‌شود. *DES* فرض می کند که داده‌ها به شکل دودویی باشند. آن را برای رمزی کردن و رمزگشایی بلوکهایی که شامل ۶۴ بیت است طرح کرده‌اند. از مقدار ۶۴ بیت کلید استفاده شده در واقع تنها از ۵۶ بیت آن استفاده می‌شود؛ ۸ بیت باقی‌مانده بررسی توازن است. از این رو تعداد کلیدها در حدود $۷,۲ \times ۱۰^{۱۶}$ می‌باشد.

دو سیستم اساسی برای رمزی کردن بلوکی ***سیستمهای رمزی انتقال*** و ***جایگزین*** می‌باشند. این روشهای رمزی کردن، تاریخ خیلی طولانی دارند. اکثر این روشها در طول

سالهای قبل معانی خود را از دست داده‌اند. آنها اغلب تا جنگ جهانی دوم بــه کــار بــرده می‌شدند، ولی از زمانی که تحلیلگران از دسترسی به رایانه سود می‌برند اغلب کمتر و کمتر به کار برده می‌شوند. با وجود این، به ایــن معنــی نیســت کــه بحــث سیســتمهای رمــزی کلاسیک تنها از نظر تاریخی مهم هستند؛ بالعکس، سیستم کلاسیک ممکن است به خــودی خود به کار برده نشود، ولی هنوز هم در سیستمهای رمزی مدرنتر به عنوان سنگ بنا مــورد استفاده قرار می‌گیرد. یک الگوریتم رمزنگاری مدرن نظیر *DES* را می‌توان به طور مثال به صورت سری از انتقالها و جایگزینی، در نظر گرفت.

یک رمز انتقالی با این حقیقت که نمادهایی که در آنها متن ساده عرضه شده تغییــر نیافته‌اند مشخّص می‌گردد، ولی تنها دنبالهٔ آنها تغییر می‌کند. در حالت رمزهای جــایگزینی دنباله تغییر نمی‌کند ولی نمادها تغییر می‌کنند؛ نمادهای اصلی در متن ســاده بــا نمادهــای دیگری جایگزین می‌شوند.

رمزهای انتقالی

در رمزهای انتقالی تنها ترتیب نمادها یا اعضایی که متــن ســاده را می‌ســازند تغیــیر می‌کنند. این کار براساس بلوک انجام می‌شود. مثال زیر را ملاحظه نمایید.

متن ساده: the invasion will begin

تقسیم به بلوکها: thein vasio nwill begin

متن رمزی: ehnti saovi iwlnl genbi

گرچه امروزه اصولاً انتقال و جایگزینی را بر بیت گروهی ترجیح می‌دهند، مثال مــا در این فصل به علّت روشنی درباره زبان است. متن ساده مثال به قســمتهایی کــه شــامل ۵ حرف است تقسیم شده‌اند. در این حالت ***طول دوره*** برابر ۵ می‌باشد. حروف بر طبق کلیــد ۳۲۵۱۴ در داخل بلوکها مجدّداً مرتب شده‌اند. از این رو نسبت به بلوک اصلی به ترتیــب حروف سوم و دوم در وضعیت اوّل و دوم، حــروف پنجــم و اوّل در وضعیتهــای ســوم و چهارم در حالی که حرف چهارم در وضعیت پنجم قرار داده شده است.

رمزگذاری پیام بر طبق رمز انتقالی را می‌توان در واقع بــه صــورت کــاربرد انتقــال ستونی تصوّر کرد، همان‌طور که از مثال زیــر روشــن خواهــد شــد. بــه جــای ایــن کــه بلوک ۵تایی از حروف را کنار یکدیگر قرار دهیم اکنون آنها را در زیــر یکدیگــر قــرار می‌دهیم.

متن ساده: the invasion will begin

کلیدواژه: کلید ۳۲۵۱۴

thein	ehnti
vasio	saovi
nwill	iwlnl
begin	genbi

متن رمزی: ehnti saovi iwlnl genbi

روشن است که پیام رمزی را می‌توان با تعویض ستونها بر طبق کلیدواژه به دست آورد.

اکنون می‌توان انتقال رمز را به صورت کلّی‌تر به طریق زیر شرح داد. برای یک طول مدّت T تعداد کلّ کلیدها $T!$، یا واقعاً $T!-1$، است، چون یک کلید وجود دارد که متن رمزی خواهد داد که با متن ساده یکی است.

در مثالی که در این بخش ارائه شد طول مدّت برابر ۵ است بدین معنا که $5!-1=119$ کلید ممکن وجود دارد، که عملاً بدون استفاده است. اگر تحلیلگر طول مدّت را بداند سریعاً قادر به رمزگشایی پیام رمزی خواهد بود.

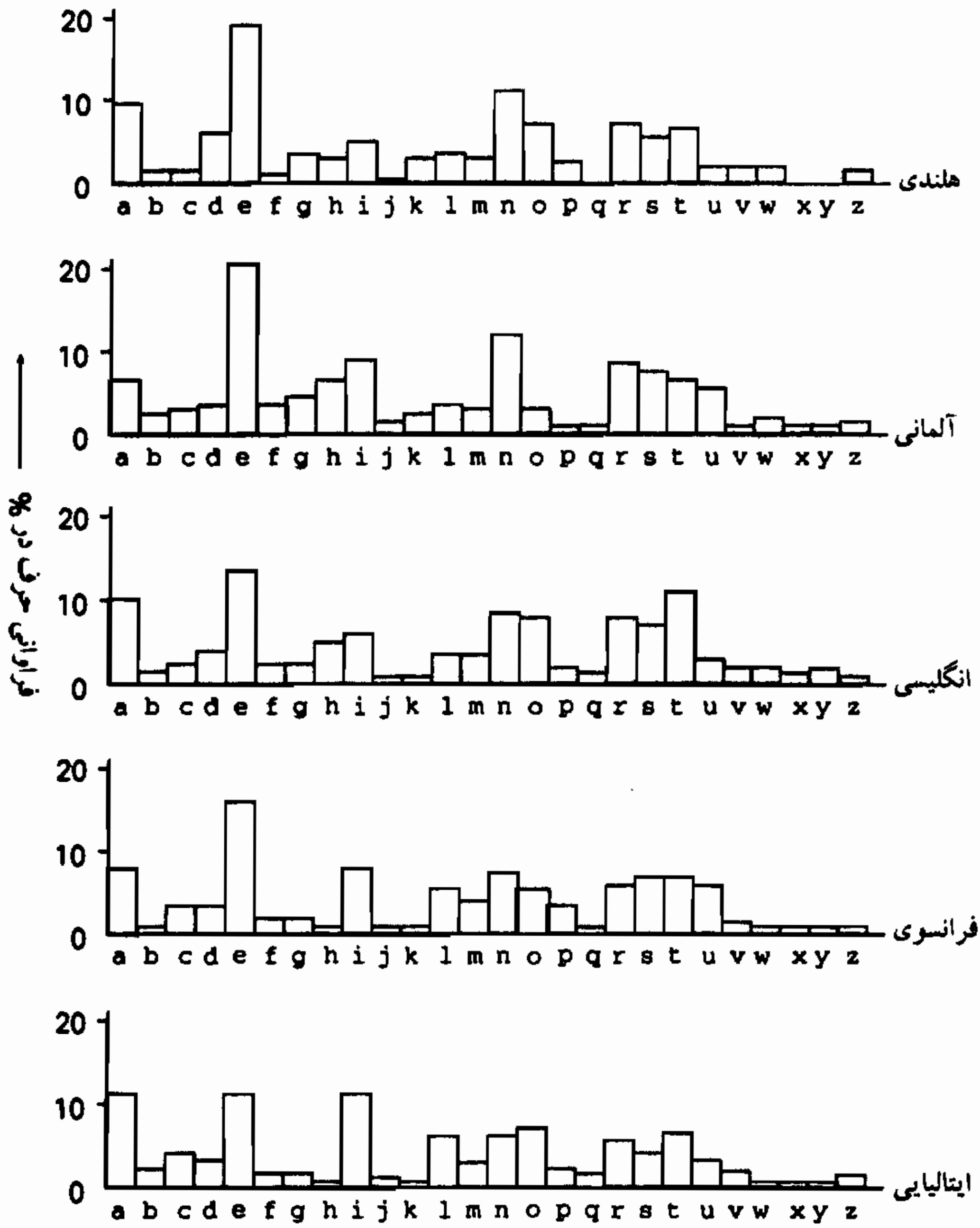
به طور کلّی تحلیلگر در رمزهای انتقالی با طول دورهٔ بزرگ با دو مسأله مواجه است. قبل از هر چیز باید بکوشد طول دوره را بیابد. در واقع برای این کار باید همهٔ اعداد n و T را که در $L=nT$ صدق می‌کنند امتحان کند، که در آن L طول پیام است. اگر این امکان که حروف ساختگی به متن ساده افزوده شده‌اند مستثنا نشده باشد، هنوز هم بایستی ترکیبهای بیشتری از n و T را امتحان کند.

مسألهٔ دوم این است که وقتی طول دوره معلوم است کلید را به روشی که ساخته شده بیابد، بدون این که مجبور باشد همه جایگشتهای ممکن را به طور کامل آزمایش کند.

در متنهای سادهٔ زبان تحلیلگر می‌تواند از مشخّصه‌های زبان برای غلبه بر این دو مسأله استفاده کند. اگر فراوانیهای حروف یک زبان را بررسی کنیم درمی‌یابیم که بعضی از حروف بیشتر از حروف دیگر ظاهر می‌شوند (شکل (۲.۱۰) را ببینید).

همین‌طور درست است اگر به فراوانیهای نسبی حروف تکی نگاه نکنیم ولی به فراوانیهای زوج حروف (دوتایی) نگاه کنیم. نتیجه می‌شود که وقتی متنی را بررسی می‌کنیم حروف صدادار اغلب توسط حروف بی‌صدا محاط شده‌اند و بالعکس. این بدین معناست که حروف صدادار به طور مساوی در سرتاسر متن گسترده خواهند بود. منظور

این است که وقتی طول دوره را تعیین می‌کنیم می‌توانیم Tای را جست‌وجو کنیم که برای آن ستونهای حاصل منظّم‌ترین توزیع حروف صدادار را روی ستونها نشان می‌دهد. در این



شکل ۱۰.۲- خارج قسمتهای فراوانی حروف برای چندین زبان مختلف

صورت بعداً در اثنای رمزگشایی با پهلوی هم قـرار دادن آن سـتونها تعـداد زیـادی زوج حرف را، که مکرراً رخ می‌دهند، خواهد داد. روشن است که این تنها وقتی ممکـن اسـت که متن ساده زبان باشد. در غیر این صورت، مشکلتر است.

رمزهای جایگزینی

در رمزهای جایگزینی، نمادهای متـن سـاده توسـط نمادهـای دیگـری جـایگزین می‌شوند. فرض کنید که الفبایی شامل ۲۶ حرف باشـد، در ایـن صـورت سیسـتم رمـزی جایگزین را می‌توان به طور کلّی به صورت زیر شرح داد

الفبا با توجّه به متن ساده: $A=[a_1,\ldots,a_{26}]$

الفبا با توجّه به متن رمزی: $B=[b_1,\ldots,b_{26}]$

متن ساده: $a_3, a_{23}, a_9, a_{17}, a_4$

متن رمزی: $b_3, b_{23}, b_9, b_{17}, b_4$

ساده‌ترین جایگزینی عبارت از جایگزینی سـزاری[1] اسـت کـه نـامش برگرفتـه از «جولیوس سزار» است. الفبای جایگزینی شامل الفبای اصلی جا به جا شـده می‌باشـد. در مثال زیر یک انتقال ۳ مکانی به کار برده شده است.

الفبای اصلی A:

a b c d e f g h i j k l m n o p q r s t u v w x y z

الفبای جایگزین B:

d e f g h i j k l m n o p q r s t u v w x y z a b c

متن ساده: the invasion will begin

متن رمزی: wkh lqydvlrq zloo ehjlq

مشخصّات جایگزینی سزاری آن است که الفبا یکسان باقی می‌ماند. تعداد کلیدها تنها بالغ بر ۲۶ می‌باشد، بنابراین متن رمزی را می‌توان به سادگی گشود. تنهـا لازم اسـت کـه حرف متناظر در متن رمزی را برای یک حرف بدانیم تا سیستم گشوده شود. اگر پیـام بـه اندازهٔ کافی بزرگ باشد چنین حرفی را به سادگی می‌توان پیدا کرد. تنهـا بـاید حرفـی را جستجو کنیم که زیادترین رخداد را در پیام رمزی دارد؛ به احتمـال زیـادی، ایـن حـرف

1. Caesar

متناظر با حرف e در متن ساده است.

اکنون اگر به جای الفبا یا جابه‌جایی حروف، الفبایی اختیار کنیم که در آن حروف به ترتیب دلخواهی قرار داده شده‌اند، در این صورت تعداد کلیدها !۲۶ (۲۶ فاکتوریل) می‌شود. این مطلب رمزگشایی را به طور قابل ملاحظه‌ای از حالت جایگزین سزاری مشکلتر می‌سازد. به یک مثال توجّه کنید

الفبای اصلی A:

a b c d e f g h i j k l m n o p q r s t u v w x y z

الفبای جایگزین B:

e s t v f u z g y x b h k w c i r j a l m p d q o n

متن ساده: the invasion will begin

متن رمزی: lgf ywpeaycw dyhh sfzyw

با وجود !۲۶ کلید ممکن، هنوز هم حلّ این‌گونه جایگزینیها نسبتاً ساده‌اند. همان‌طور که می‌دانیم زبان با درجهٔ بالایی مازاد بر احتیاج است. علاوه بر این، همواره می‌توان حروف نظیر e، t، a، n و غیره را که زیادترین ظهور را دارند براساس توزیع فراوانی حروف در متن رمزی به سادگی به دست آورد.

نتیجه می‌شود که روشهای جایگزینی همان‌طور که در بالا شرح داده شد نیرومند نیستند. این بدین علّت است که ویژگیهای زبان هنوز هم نسبتاً ساده‌اند که از متن رمزی استخراج کنیم. بنابراین اغلب بیش از یک جایگزینی به کار برده می‌شود. در این حالت ***جایگزینی چند الفبایی*** داریم. یک مثال مشهور جایگزین چند الفبایی سیستم ویگنر[1] است. از این نتیجه می‌شود که نه یک بلکه تعدادی جایگزین سزار به کار برده می‌شوّد. برای مثال، اوّلین حرف در متن ساده بیش از ۲۰ مکان و حرف دوم بیش از ۱۷ مکان جا به جا شده است و الی آخر.

اغلب از جدولی موسوم به جدول ویگنر (شکل (۳.۱۰) را ببینید) و کلید واژهٔ انتخابی استفاده می‌شود. در جدول ویگنر حروف الفبای متن ساده در ردیف بالا قرار داده شده‌اند. ستون اوّل شامل حروف ممکن کلیدواژه است. اکنون رمزگشایی به صورت زیر انجام می‌شود.

کلیدواژه در زیر متن ساده چنان که در مثال زیر نشان داده شده است قرار می‌گیرد.

---
1. Vignére

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a
c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b
d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c
e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d
f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e
g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f
h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g
i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h
j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i
k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j
l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k
m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l
n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m
o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n
p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q
s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
u	v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
v	w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u
w	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v
x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x
z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y

شکل ۳.۱۰- جدول ویگنر

اکنون حروف پیام رمزی با انتخاب حرفی از جدول که در تقاطع ستون با حرف متن ساده و سطر با حرف کلیدواژه قرار دارد به دست می‌آید.

متن ساده: the invasion will begin

کلید: rad ioradior adio radio

متن رمزی: khh qbmavqce wltz sejqb

نتیجه این است که یک حرف در متن ساده می‌تواند بـا حـروف متفـاوتی در متـن رمزی در ارتباط با حرف کلیدواژه نشان داده شود. به این علّت، مشخّصات زبـان بهـتر از روشهای مذکور قبلی مخفی نگه‌داشته می‌شود.

تعداد جایگزینهای تک الفبایی که اساس سیستم ویگنر را تشکیل می‌دهد برابر طـول کلیدواژه است. برای مثال داده شده در بالا تعداد جایگزین تک الفبایی به کار بـرده شـده بالغ بر ۵ می‌باشد؛ یعنی پنج سطر از جدول به کار برده شده است.

روشن است که آگاهی از طول کلیدواژه کمک بزرگی بـه تحلیلگـر در حـلّ پیـام رمزی است.

کاربر سیستم معمولاً برای حدّاکثر استفاده از سطرهای جدول در اثنـای رمزگشـایی تلاش خواهد کرد. یک روش رسیدن به آن، این است که از خود متـن سـاده و همچنیـن کلیدواژه همان‌طور که در زیر شرح داده شده استفاده شود

متن ساده:	the invasion will begin
کلید:	rad iotheinv asio nwill
متن رمزی:	khh qbohwqbi watz oaoty

پس از به کار بردن کلید « radio » متن ساده خود نیز با عنوان کلید اسـتفاده شـده است.

آخرین تذکر؛ مسألهٔ کاربرد انتقال و جایگزینی رمزها آن است که مشخّصات متـن سادهٔ اصلی تا جای ممکن پنهان نگه‌داشته شود. یکی از راه‌حلها این است که مطمئن شـویم که حروف یا نمادها در متن رمزی دارای توزیع یکنواخت اسـت. قبـل از بـه کـارگیری روش انتقال یا جایگزینی، با رمزگشایی حروف متن ساده و با به کار بردن روشـی نظـیر هافمن، همان‌طور که قبلاً به آن اشاره شد، می‌تـوان بـه آن دسـت یـافت. چنانچـه همـهٔ کدنمادها دارای احتمال رخ‌داد یکسان باشند یک کدگذاری منبع بهینه حاصل خواهد شد.

۴.۱۰ مقدار اطّلاع و اطمینان

برای بهره‌وری از سیستم رمزنگاری داشتن احساس اطمینان از سیستم بـه کـار بـرده شده اهمیت زیادی دارد. با استفاده از مفاهیم نظریهٔ اطلاع می‌کوشیم در مورد این سؤال که واقعاً سیستم رمزی مطمئن چیست شناختی پیدا کنیم.

با در نظر گرفتن پیامهایی در متن ساده به عنوان اعضای مجموعهٔ پیامهای ممکن تولید شده توسط منبع، هر یک با احتمال رخداد خودش، می‌توان مقدار اطلاع در متن ساده را بیان کرد.

مقدار اطلاع در متن ساده به صورت زیر فرمول‌بندی می‌شود

$$H(M)=-\sum_{i=1}^{n}p(M_i)\log p(M_i), \qquad (۱.۱۰)$$

که در آن $p(M_i)$، $i=1,\dots,n$، احتمال رخداد پیامهای متن ساده M_i می‌باشند. می‌توان مقدار اطلاع متن رمزی که با $H(C)$ نشان داده می‌شود و مقدار اطلاع $H(K)$ مربوط به کلیدها را به همین روش بیان کرد.

در روشی مشابه، می‌توان مقدار اطلاع شرطی را نیز بیان کرد. اگر دربارهٔ متن رمزی C آگاهی داشته باشیم، که به ***ابهام کلید*** نیز مشهور است، در این صورت $H(K|C)$ مقدار اطلاع یا عدم‌حتمیت با توجّه به کلید است. آن را می‌توان به صورت زیر تعریف کرد. فرض کنید K_h، $h=1,\dots,l$، مجموعهٔ کلیدها باشند و C_j، $j=1,\dots,m$، پیامهای رمزی ممکن باشند، در این صورت داریم

$$H(K|C)=-\sum_{h=1}^{l}\sum_{j=1}^{m}p(K_h,C_j)\log p(K_h|C_j). \qquad (۲.۱۰)$$

به طور مشابه، $H(M|C)$ مقدار اطلاع یا عدم‌حتمیت مربوط به متن سادهٔ M برای یک متن رمزی داده شده است که آن را ***ابهام پیام*** نیز می‌نامند. $H(M|C,K)$ را می‌توان به طور مشابه مقدار اطلاع متن ساده وقتی هر دو متن رمزی و کلید معلومند در نظر گرفت.

چون متن ساده توسط متن رمزی و کلید، بدون ابهام تعیین می‌شود، داریم

$$H(M|C,K)=0. \qquad (۳.۱۰)$$

اگر دسترسی به متن رمزی و کلید داشته باشیم، ممکن است متن ساده را نیز تعیین کرد. در این صورت عدم‌حتمیت دربارهٔ M برابر صفر است.

کمیّت $H(K|M,C)$، موسوم به ***ابهام حضور کلید***، مقدار اطلاع مربوط به کلید با معلوم بودن متن ساده و متن رمزی می‌باشد.

قضیهٔ ۱.۱۰

برابری زیر برای ابهام حضور کلید برقرار می‌باشد:

$$H(K|M,C)=H(K|C)-H(M|C). \qquad (۴.۱۰)$$

برهان

براساس رابطه‌های داده شدهٔ قبلی برای مقدار اطلاع توأم، مقدار اطلاع در متن ساده، متن رمزی و کلید را به دست می‌آوریم که برابر است با

$$H(M,C,K)=H(M|C,K)+H(C,K)$$
$$=H(K|M,C)+H(M,C). \qquad (۵.۱۰)$$

به خاطر بیاورید که همچنین داریم

$$H(C,K)=H(K|C)+H(C),$$

و

$$H(M,C)=H(M|C)+H(C),$$

در این صورت از معادلهٔ (۵.۱۰) نتیجه می‌شود که

$$H(M|C,K)+H(K|C)=H(K|M,C)+H(M|C). \qquad (۶.۱۰)$$

قبلاً در بالا دیدیم که $H(M|C,K)=0$ (معادلهٔ (۳.۱۰) را ببینید). در این صورت معادلهٔ (۴.۱۰) با کمک معادلهٔ (۶.۱۰) به دست می‌آید. ■

قضیهٔ (۱.۱۰) به برخی تفاسیر جالبی منجر می‌شود. از نقطه‌نظر کاربر برای مقدار بزرگی از $H(K|M,C)$ تلاش می‌شود. اگر تحلیلگر هر دو متن ساده و رمزی را در دسترس داشته باشد، در این صورت در هر حالتی باید مطمئن شود که عدم‌حتمیت دربارهٔ کدام کلیدی که استفاده شده به قدر ممکن بزرگ است. براساس معادلهٔ (۴.۱۰) می‌توان نتیجه گرفت مقدار بزرگی برای $H(K|M,C)$ را می‌توان به دست آورد با اطمینان این که $H(M|C)$ مقدار کوچکی اختیار کند. با وجود این، یک مقدار کوچک $H(M|C)$ به این معناست که عدم‌حتمیت کوچکی دربارهٔ متن ساده M وجود دارد اگر تنها به متن رمزی C دسترسی داشته باشیم. با وجود این، به بیان دقیقتر می‌خواهیم از این اجتناب کنیم. در واقع به این علّت با تنگنایی مواجه می‌شویم. عدم‌حتمیت زیاد در ارتباط با کلید به بهای عدم‌حتمیت دربارهٔ متن سادهٔ ارسال شده روی می‌دهد. بالعکس، یک عدم‌حتمیت بزرگ در ارتباط با متن ساده با عدم‌حتمیت کوچکی در ارتباط با کلید همراه است.

با کمک بررسی این اطلاعات نظری نتایج مهم دیگری می‌توان به دست آورد. فرض کنید $I(M;C)$ اطلاع متقابل بین متن ساده و متن رمزی باشد که به صورت زیر تعریف می‌شود

$$
\begin{aligned}
I(M;C) &= H(M) - H(M|C) \\
&= H(C) - H(C|M). \qquad (۷.۱۰)
\end{aligned}
$$

از نقطه‌نظر کاربر سیستم رمزی تلاش خواهد شد که $I(M;C)$ تا حدّ ممکن کوچک شود، چون اطلاع متقابل اندازهٔ استقلال متقابل است و این باید در این حالت کوچک باشد. اگر متن رمزی مطلقاً هیچ اطلاعی دربارهٔ متن ساده ندهد، در این صورت $H(M|C) = H(M)$. بنابراین از معادلهٔ (۷.۱۰) نتیجه می‌شود که اطلاع متقابل بین متن ساده و متن رمزی برابر صفر خواهد شد.

یک ***سیستم رمزی مطلقاً مطمئن*** داریم اگر

$$I(M;C) = ۰. \qquad (۸.۱۰)$$

رابطهٔ مهمی برای سیستمهای رمزگذاری در قضیهٔ زیر ارائه شده است.

قضیهٔ ۲.۱۰

فرض کنید $I(M;C)$ اطلاع متقابل بین متن ساده و متن رمزی باشد، در این صورت داریم:

$$I(M;C) \geq H(M) - H(K). \qquad (۹.۱۰)$$

برهان

برای اثبات این رابطه معادلهٔ (۴.۱۰) را مجدداً بررسی می کنیم. چون $H(K|M,C) \geq ۰$، از معادلهٔ (۴.۱۰) می‌توان به دست آورد که

$$H(K|C) \geq H(M|C). \qquad (۱۰.۱۰)$$

و بنابر تعریف

$$H(K) \geq H(K|C),$$

از فرمول (۱۰.۱۰) نتیجه می‌شود که

$$H(K) \geq H(M|C). \qquad (۱۱.۱۰)$$

اگر این معادله را با معادلهٔ (۷.۱۰) ترکیب کنیم در این صورت فرمول (۹.۱۰) بی‌درنگ نتیجه می‌شود. ■

رابطهٔ فرمول (۹.۱۰) واقعاً بیان می کند که مجموعه‌ای از کلیدها که شامل اطلاع

کمّی (به طور متوسط) است یک اطلاع متقابل بزرگ ممکن بین متن ساده و متن رمزی می‌سازد.

اطمینان مطلق، یعنی $I(M;C)=0$، تنها در حالتی به دست می‌آید که

$$H(K) \geq H(M). \qquad (۱۲.۱۰)$$

بنابراین اطلاع در کلید حداقل بایستی به اندازهٔ اطلاع در متن ساده باشد.

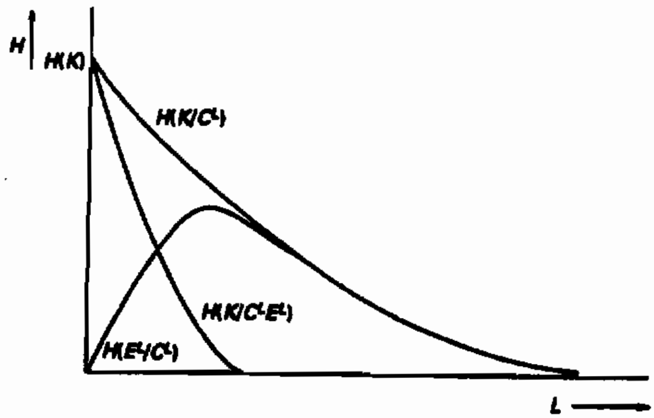
۵.۱۰ فاصلهٔ یکتایی

در بخش قبل طول متن رمزی جدا شده توسط تحلیلگر منظور نشده است. با این حال از اهمیت خاصی برخوردار است. فرض کنید که متن ساده مربوط به زبان است. همان‌طور که قبلاً یادآوری شد زبان برخی ویژگیهای آماری را نشان می‌دهد. براساس این ویژگیهای آماری واقعی اغلب برای تحلیلگر امکان‌پذیر است که بخشی یا همهٔ متن رمزی را رمزگشایی کند.

به طور کلّی درست است که هر چه متن رمزی بزرگتری در اختیار تحلیلگر باشد، او با احتمال بیشتری قادر خواهد بود که کلید را به دست آورد. اگر متن رمزی به طول L را به صورت C^L نشان دهیم، در این صورت ابهام کلید $H(K|C^L)$ برای مقادیر زیاد L کاهش خواهد یافت، زیرا اگر L افزایش یابد عدم‌حتمیت دربارهٔ کلید به کار برده شده کاهش می‌یابد. این مطلب در شکل(۴.۱۰) شرح داده شده است. ابهام کلید حتّی در زمان داده شده‌ای صفر خواهد شد، وقتی کلید را بتوان از متن رمزی با اطمینان پیدا کرد. ماکسیمم مقدار ابهام کلید برابر $H(K)$ خواهد بود. این مقدار وقتی احتمالهای یک کلید به شرط متن رمزی برابر با احتمال خود کلید باشد، رخ خواهد داد. در این صورت در ارتباط با کلید استفاده شده عدم‌حتمیت داریم.

بررسی مشابهی برای ابهام پیام $H(M^L|C^L)$ به کار می‌رود. با وجود این، یک اختلاف وجود دارد: تعداد کلّ پیامها هنوز هم برای مقادیر کوچک L کوچک خواهد بود، بنابراین به عنوان نتیجهٔ ابهام پیام نیز مقدار کوچکی اختیار خواهد کرد. با وجود این، اگر L افزایش یابد تعداد پیامهای ممکن سریعاً افزایش خواهد یافت و بنابراین ابهام پیام نیز افزایش می‌یابد اگر L مقداری به قدر کافی بزرگ اختیار کند. در این صورت متن رمزی یا پیام رمزی سرانجام شامل اطلاع کافی خواهد بود که مانع افزایش بیشتر تعداد محتملترین پیامها بشود. اندکی پس از این نقطه حتی ابهام پیام منطبق بر ابهام کلید خواهد

بود، چون در این صورت پیام رمزی شامل همهٔ اطلاع برای تعیین کلید از متن سادهٔ یافته شده با همان اطمینان می‌باشد و بالعکس.



شکل ۴.۱۰- کلید، پیام و ابهام حضور کلید به عنوان تابعی از L

مسیر ابهام حضور کلید $H(K|C^L, M^L)$ نیز در یک شکل رسم شده است. به طور وضوح ابهام حضور کلید سریعتر از ابهام کلید به صفر میل می‌کند، زیرا در حالت ابهام حضور کلید فرض شده است که تحلیلگر متن ساده و متن رمزی را نیز دارد. این آگاهی اضافی موجب خواهد شد که تحلیلگر روی‌هم‌رفته کلید را سریعتر به دست آورد.

ابهام حضور کلید اندازه‌ای برای پایداری سیستم پیام رمزی تحت اقدام متن ساده-معلوم برای کلید است، در حالی که ابهام کلید و پیام اندازه‌ای برای پایداری سیستم پیام رمزی تحت اقدام تنها با متن رمزی به ترتیب برای کلید و پیام می‌باشد.

از مطلب بالا روشن خواهد شد که همان‌طور که طول متن رمزی دریافت شده بزرگتر می‌شود، احتمال این که تحلیلگر قادر باشد کلید یا متن ساده را بیابد افزایش می‌یابد.

قضیهٔ ۳.۱۰

فرض کنید ε تعداد نمادهای مختلف در یک پیام یا متن رمزی به طول L باشد. برای ابهام کلید داریم

$$H(K|C^L) \geq H(K) - D_L, \qquad (۱۳.۱۰)$$

که در آن D_L حشو مطلق نامیده می‌شود که به صورت زیر تعریف می‌شود

$$D_L = L\log(\varepsilon) - H(M^L). \tag{۱۴.۱۰}$$

برهان

چون یک رابطهٔ بدون ابهام بین متن ساده و متن رمزی وجود دارد همواره خواهیم داشت که

$$H(K,C^L) = H(K,M^L).$$

تحت این شرط که کلید از پیام منبع مستقل است، بنابراین برای ابهام کلید داریم

$$\begin{aligned} H(K\,|\,C^L) &= H(K,C^L) - H(C^L) \\ &= H(K,M^L) - H(C^L) \\ &= H(K) + H(M^L) - H(C^L). \end{aligned} \tag{۱۵.۱۰}$$

چون ε تعداد نمادهای مختلف یک پیام یا متن رمزی است، ε^L پیام ممکن یا پیام رمزی به طول L وجود دارند. براساس ویژگیهای اندازهٔ اطلاع می‌توان ثابت کرد که

$$H(C^L) \le L\log(\varepsilon).$$

با این رابطه و با کمک معادلهٔ (۱۵.۱۰) نتیجه می‌شود

$$H(K\,|\,C^L) \ge H(K) + H(M^L) - L\log(\varepsilon). \tag{۱۶.۱۰}$$

جملهٔ $[H(M^L) - L\log(\varepsilon)]$ در معادلهٔ (۱۶.۱۰) مقدار منفی حشو D_L است، بنابراین معادلهٔ (۱۶.۱۰) دقیقاً معادلهٔ (۱۳.۱۰) می‌باشد. ■

حشو مطلق را می‌توان به صورت اندازهٔ وسعت اختلافی که منبع واقعی از منبعی، که در آن هر پیام با احتمال برابر رخ می‌دهد، دارد، در نظر گرفت. این دقیقاً وابسته به تعریف حشو در معادلهٔ (۳.۲) می‌باشد که در واقع اندازهٔ حشو نسبی است.

یک تفسیر نابرابری داده شده در قضیهٔ (۳.۱۰) آن است که اگر حشو افزایش یابد ابهام کلید روی هم رفته کاهش می‌یابد و بنابراین عدم‌حتمیت دربارهٔ کلید نیز کاهش می‌یابد. به عبارت دیگر، روش تقلیل حشو از قرار معلوم اطمینان سیستم رمزی را افزایش می‌دهد. تفسیر دیگر این است که تا زمانی که $H(K) > D_L$ ابهام کلید $H(K\,|\,C^L)$ نمی‌تواند برابر صفر شود، بنابراین روی هم رفته ممکن نیست کلید را به صراحت پیدا کرد.

اگر طول پیام رمزی کوچک باشد این مطلب رخ می‌دهد. این مطلب با فرض این که

منبع پیام بی‌حافظه است و در رابطه زیر صدق می‌کند مشخص می‌شود

$$H(M^L) = LH(M). \qquad (۱۷.۱۰)$$

یعنی، مقدار اطلاع در هر پیام L برابر مقدار اطلاع در هر نماد می‌باشد.

اگر معادلهٔ (۱۷.۱۰) را در فرمول (۱۳.۱۰) جایگزین کنیم در این صورت داریم

$$H(K|C^L) \geq H(K) + L[H(M) - \log(\varepsilon)]. \qquad (۱۸.۱۰)$$

از این رو ابهام کلید می‌تواند روی‌هم‌رفته صفر شود اگر

$$L \geq \frac{H(K)}{\{\log(\varepsilon) - H(M)\}}. \qquad (۱۹.۱۰)$$

به عبارت دیگر، اگر مقدار اطلاع منبع، یعنی عدم‌حتمیت، کوچک باشد تنها چند نماد برای یافتن کلید لازم است. بنابراین، در عمل، برای مثال با به کار بردن کدگذاری منبع کاربر باید مطمئن باشد که اطلاع پیامها به قدر ممکن بزرگ است. مقدار L که برای آن در فرمول (۱۹.۱۰) برابری برقرار است ***فاصلهٔ یکتایی*** نامیده می‌شود و با UD نمایش می‌دهند. برای این که قادر به یافتن کلید باشیم می‌نیمم طول متن رمزی لازم است.

باید یادآوری شود که فاصلهٔ یکتایی نشان می‌دهد که مقداری برای L وجود دارد که به موجب آن $H(K|C^L)$ می‌تواند برابر صفر باشد. این بدین معنا نیست که در این صورت کلید را می‌توان یافت. در این جا، مانند بقیهٔ فصل، رهیافت براساس حالت متوسط است. در کاربرد مفاهیم در نظریهٔ اطلاع توجّه به متوسط مقدار اطلاع، امری ذاتی است.

۶.۱۰ تمرینها

۱.۱۰ یک سیستم رمزی جایگزین کلّی با کلیدهای هم‌احتمال را در نظر بگیرید که در آن پیامهای M^L به طول L به عنوان متن ساده مشخّص شده‌اند و با کمک کلید K به متن رمزی C انتقال داده می‌شوند.

فرض کنید که منبع بی‌حافظه باشد و نمادها از الفبای $U = (u_1, u_2, \ldots, u_8)$ با احتمال رخ‌داد زیر تولید می‌شوند:

$$P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{128})$$

(الف) مقدار اطلاع در متن ساده را تعیین کنید. همچنین مقدار اطلاع در متن رمزی را

بیابید.
(ب) مقدار اطلاع در کلید را به دست آورید.
(پ) فاصلهٔ یکتایی را بیابید.
(ت) فاصلهٔ یکتایی را وقتی جایگزین سزار به جای جایگزین کلّی به کار برده شده است تعیین کنید.

۲.۱۰ یک سیستم رمزی کلّی را در نظر بگیرید. در فرستنده یک متن ساده M توسط تبدیل T، که وابسته به کلید K است، به متن رمزی C ارسال شده است.
فرض کنید که منبع اطلاع بی‌حافظه‌ای که متنهای ساده‌ای شامل نمادهایی از الفبای $V = (v_۱, v_۲, v_۳, v_۴)$ را تولید می‌کند به کار برده است. احتمالهای نمادها به صورت زیر داده شده‌اند

$$p_۱ = \frac{۷}{۱۶} \quad , \quad p_۲ = \frac{۳}{۱۶} \quad , \quad p_۳ = \frac{۲}{۱۶} \quad , \quad p_۴ = \frac{۴}{۱۶}$$

برای رمزگذاری یک جایگزین تک‌الفبایی به کار برده شده است. فرض شده است که هر الفبای جایگزین دارای احتمال رخداد یکسان است.

(الف) مقدار حشو مطلق در متن ساده و مقدار اطلاع را در کلید محاسبه کنید.
(ب) فاصلهٔ یکتایی برای این حالت را محاسبه کنید و تفسیری از نتیجه ارائه نمایید.
اکنون فرض کنید برای رمزگذاری متن ساده دو نماد متوالی متن ساده به عنوان یک نماد جدید در نظر گرفته می‌شود. الفبای جایگزین عبارت است از $U = (u_۱, u_۲, u_۳, \dots)$.
(پ) ارزیابی اندازهٔ این الفبای جایگزین را باید داشته باشیم. هر مقدار اطلاع در کلید و حشو متن ساده را محاسبه کنید.
(ت) فاصلهٔ یکتایی را در این حالت محاسبه کنید و اختلاف با فاصلهٔ یکتایی محاسبه شده در (ب) را شرح دهید.

۳.۱۰ فرض کنید یک منبع متنهای ساده‌ای شامل حروفی از الفبای طبیعی مشتمل بر ۲۶ حرف را تولید می‌کند. فرض شده است که مقدار اطلاع در متن ساده همانند آنتروپی زبان انگلیسی است (فرض کنید ۱٫۵ بیت بر ثانیه). برای تولید متن رمزی یک جایگزین چند الفبایی با یک جدول ویگنر با طول کلید ۷ حرف به کار برده شده است.

(الف) مقدار اطلاع در کلید را از نقطه نظر تحلیلگر در حالتی که او می‌داند که کلیدواژه شامل ۷ حرف است ولی نمی‌داند کدام حروفند محاسبه کنید.
(ب) قسمت (الف) را اکنون با این فرض که ۷ حرف نباید متفاوت باشند، انجام دهید.
(پ) فاصلهٔ یکتایی را در حالت (الف) محاسبه کنید.
(ت) همچنین فاصلهٔ یکتایی را در حالت (ب) محاسبه کنید. اختلاف را شرح دهید.

۷.۱۰ جوابها

۱.۱۰ (الف) چون منبع بی‌حافظه است و از این رو $H(M^L) = LH(M) = LH(U)$، کافی است مقدار اطلاع را در سطح نماد بررسی کنیم. برای متن ساده به دست می‌آوریم

$$H(M) = H(U) = -\sum_{i=۱}^{۸} p(u_i)\log p(u_i) \text{ نماد / بیت.}$$

$$= -\frac{۱}{۲}\log\frac{۱}{۲} - \frac{۱}{۴}\log\frac{۱}{۴} - \frac{۱}{۸}\log\frac{۱}{۸} - \frac{۱}{۱۶}\log\frac{۱}{۱۶} - \frac{۱}{۳۲}\log\frac{۱}{۳۲}$$

$$-\frac{۱}{۶۴}\log\frac{۱}{۶۴} - ۲\times\frac{۱}{۱۲۸}\log\frac{۱}{۱۲۸} = ۱{,}۹۸ \text{ نماد / بیت.}$$

چون برای جایگزین رمزی یک تناظر یک به یک بین نمادهای متن ساده و نمادهای متن رمزی وجود دارد برای مقدار اطلاع در متن رمزی داریم که

$$H(C) = H(M) = ۱{,}۹۸ \text{ نماد / بیت.}$$

(ب) تعداد کلیدها !۸ است (به طور دقیقتر ۱−!۸، به استثنای کلیدی که منجر به برابری متن رمزی با متن ساده می‌شود). اطلاع کلید نتیجه می‌دهد

$$H(K) = \log ۸! = ۱۵{,}۲۹ \text{ بیت.}$$

(پ) فاصلهٔ یکتایی برابر است با

$$UD = \frac{H(K)}{\log(\varepsilon) - H(M)}$$

که در آن ε حجم الفبای منبع است.

در نتیجه

$$UD = \frac{۱۵٫۲۹}{\log ۸ - ۱٫۹۸} \approx ۱۵.$$

این متوسط می‌نیمم تعداد نمادهای متن رمزی لازم برای یافتن کلید از متن رمزی است.

(ت) تعداد کلیدها برای جایگزین سزاری برابر ۸ است. اکنون، فاصلهٔ یکتایی کاهش می‌یابد به

$$UD = \frac{\log ۸}{\log ۸ - ۱٫۹۸} \approx ۳.$$

۲.۱۰ (الف) برای اطلاع از متن ساده داریم که

$$H(M) = H(V) = -\frac{۷}{۱۶}\log\frac{۷}{۱۶} - \frac{۳}{۱۶}\log\frac{۳}{۱۶} - \frac{۲}{۱۶}\log\frac{۲}{۱۶} - \frac{۴}{۱۶}\log\frac{۴}{۱۶}$$

$$\approx ۱٫۸۵ \text{ بیت / نماد}.$$

حشو مطلق برابر است با

$$D = \log(\varepsilon) - H(M) = \log ۴ - ۱٫۸۵ \approx ۰٫۱۵ \text{ بیت / نماد}.$$

تعداد کلیدها !۴ (یا ۱–!۴) است و از این رو چون آنها هم‌احتمالند داریم

$$H(K) = \log ۴! = \log ۲۴ \approx ۴٫۵۸ \text{ بیت}.$$

(ب) فاصلهٔ یکتایی برابر است با

$$UD = \frac{H(K)}{\log(\varepsilon) - H(M)} = \frac{۴٫۵۸}{۰٫۱۵} \approx ۳۱.$$

(پ) چون الفبای اصلی از چهار نماد وجود دارد، با ترکیب دو نماد منجر به ۱۶ نماد ترکیب شدهٔ جدید می‌شود.

اکنون، حجم الفبای جایگزین باید ۱۶ باشد و از این رو برای مقدار اطلاع در کلید به دست می‌آوریم

$$H(K) = \log ۱۶! = ۴۴٫۳ \text{ بیت}.$$

به طور کلّی، حشو مطلق برای پیامها به طول L در حالت منبع بی‌حافظه برابر است با

$$D_L = L\log(\varepsilon) - H(M^L).$$

در حالت موجود به دست می‌آوریم

$$D_۲ = ۲\{\log - H(M)\} = ۲D = ۰٫۳۰ \text{ پیام / بیت}.$$

با ترکیب زوج پیامها منجر به فاصلهٔ یکتایی می‌شود

$$UD = \frac{H(K)}{D_L} = \frac{۴۴٫۳}{۰٫۳۰} = ۱۴۸.$$

برای این که این را با نتیجهٔ قسمت (ب) مقایسه کنیم باید این فاصلهٔ یکتایی را بر ۲ تقسیم کنیم که ۷۴ نتیجه می‌شود، چون اکنون طول یک نماد به اندازهٔ دو برابر طول در حالت (ب) می‌باشد. روشن است که حشو در هر نماد تغییر نمی‌کند تعداد کلیدها افزایش یافته است.

۳.۱۰ (الف) چون کلید شامل ۷ حرف متفاوت است، تعداد کلیدها عبارتند از

$$۲۶ \times ۲۵ \times ۲۴ \times ۲۳ \times ۲۲ \times ۲۱ \times ۲۰ = \frac{۲۱!}{۱۹!}$$

و بنابراین مقدار اطلاع کلید برابر است با

$$H(K) = \log \frac{۲۶!}{۱۹!} \approx ۳۱٫۶۳ \text{ بیت}.$$

(ب) در حالت کلی به دست می‌آوریم

$$H(K) = \log ۲۶^۷ = ۳۲٫۹۰ \text{ بیت}$$

(پ) برای فاصلهٔ یکتایی به دست می‌آوریم

$$UD = \frac{H(k)}{\log(\varepsilon) - H(M)} = \frac{H(K)}{\log ۲۶ - ۱٫۵} = \frac{H(K)}{۳٫۲}.$$

(ت) در حالت (الف) به دست می‌آوریم $UD_a = \frac{۳۱٫۶۳}{۳٫۲۰} \approx ۱۰$ و در حالت (ب) به دست می‌آوریم: $UD_b = \frac{۳۲٫۹۰}{۳٫۲۰} \approx ۱۱$. چون فضای کلید (الف) کوچکتر از فضای (ب) است، به طور کلی نماد کمتری برای یافتن کلید لازم است. از این رو داریم $UD_a \le UD_b$.

کتاب‌نامه

Abramson, N. (1963), *Information theory and coding*, McGraw-hill Book Company, New York.

Aczel, J. and Z. Daroczy (1975), *On measure of information and their characterization*, Academic Press, New York.

Ash, R.B. (1965), *Information theory*, Interscience, New York.

Azimoto, S. (1971), Information theoretical considerations on estimation problems, *Inform. contr*. Vol. 19, pp. 181-194.

Bell, D.A. (1962), *Information theory and its engineering applications*, I. Pitman Ltd, London.

Berger, T. (1971), *Rate distortion theory: a mathematical basis for data compression*, Prentice-Hall, Englewood Cliffs, NJ.

Blahut, R.E. (1987), *Principles and practice of information theory*, Addison-Wesley, Reading, Mass.

Blahut, R.E. (1990), *Digital transmission of information*, Addison-Wesley, Reading, Mass.

Boekee, D.E. and J.C.A van der Lubble (1988), *Informatiethrorie*, Delftse Uitgevers Maatschappij, Delft.

Chaundy, T.W. and McLeod, J.B. (1960), On a functional equation, *proc. Edinburgh Math. Soc. Notes*, 43, pp. 7-8.

Cover, Th. M., and J.A. Thomas (1991), *Elements of information theory*, Wiley, New York.

Csiszár, I., Körner, J. (1981), *Information theory*, Academic Press, New York.

Daroczy, Z. (1970), Generalized information functions, *Inform. Contr.*, Vol. 16, pp. 36-51.

Fano, R.M. (1961), *Transmission of information: a statistical theory of communication*, Wiley, New York.

Feinstein, A. (1958), *Foundations of information theory*, McGraw-Hill, New York.

Feller, W. (1957), *An introduction to probability theory and its applications*, Wiley, New York.

Gallager, R.G (1968), *Information theory and reliable communication*, Wiley, New York.

Goldman, S. (1953), *Information theory*, Perntice-Hall, Englewood Cliffs, NJ.

Guiasu, S. (1976), *Information theory with applications*, McGraw-Hill, New York.

Hartley, R.V.L. (1928), Transmission of information, *Bell Syst. Tech. J.*, Vol. 7, pp. 535-563.

Jelinek, F. (1968), *probabilistic information theory*, McGraw-Hill, New York.

Lubbe J.C.A. van der (1981), *A generalized probabilistic theory of the measurement of certainty and information* (PhD thesis), Delft University of Technology, Dept. E.E., Information Theory Group.

McEliece, R.J. (1977), *The theory of information and coding*, Addison-Wesley, Reading, Mass.

McMillan, B. (1953), The basic theorems of information theory, *Ann. Math. Statist.*, pp. 196-219.

Nyquist, H. (1924), Certainty factors affecting telegraph speed, *Bell Syst. Tech. J.*, Vol. 3, pp. 324-346.

Renyi, A. (1960), On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Statist. and Prob.*, no. 1, pp. 547-561.

Shannon, C.E. (1948), The mathematical theory of communication, *Bell Syst. Tech. J.*, Vol. 27, pp. 379-423 and pp. 623-656.

واژه‌نامه

absolute security	اطمینان مطلق
Ackoff	آکوف
additive channels	کانالهای جمعی
additivity	جمع پذیری
– information measure	- اندازهٔ اطلاع
alphabet extension	توسعه الفبا
alphabetic code	کد الفبایی
American traditions	سُنت امریکایی
amount of information	مقدار اطلاع
– maximum	- ماکسیمم (بیشینه)
– minimum	- می‌نیمم (کمینه)
arithmetic code	کد حسابی
autocorrelation function	تابع خودهمبستگی
autocorrelation matrix	ماتریس خودهمبستگی
autocovariance	اتوکوواریانس
autocovariance matrix	ماتریس اتوکوواریانس
average code word length	متوسط طول کدواژه
average distortion	متوسط دگرشکلی
axiomatic foundations	اصول موضوعه
Azimoto	آزیموتو
Bar – Hillel	بار-هیلل
Bayes' theorem	قضیهٔ بیز
BEC, see binary erasure channel	BEC، رجوع کنید به کانال پاک شدگی دودویی
binary channel	کانال دودویی
binary erasure channel	کانال پاک شدگی دودویی
binary informatin source	منبع اطلاع دودویی

binary memoryless information source	منبع اطلاع بی‌حافظهٔ دودویی
binary multiplying channel	کانال چندگانه دودویی
binary symmetric channel	کانال متقارن دودویی
bit	بیت
block code, linear	کد بلوکی، خطی
block enciphering	رمزی‌کردن بلوکی
British traditions	سنتهای انگلیسی
broadcast channel	کانال پخش
– gaussian	- گاوسی
burst errors	خطاهای ناگهانی
Caesar substitution	جایگزینی سزار
capacity	ظرفیت
– C of a discrete noiseless channel	- ظرفیت کانال بدون نوفه گسسته
– of a continuous channel	- کانال پیوسته
– of noiseless channels	- کانال بدون نوفه
– of noisy channels	کانالهای نوفه‌دار
– region	- ناحیه
capacity bounds	کرانهای ظرفیت
carnap	کارنَپ
cascading	متوالی
cascading of channels	کانالهای متوالی
channel	کانال
– additive	- جمعی
– binary multiplying	- چندگانه دودویی
– broadcast	- پخش
– capacity of noiseless	- ظرفیت بدون نوفه
– cascading	- متوالی
– gaussian with memory	- گاوسی باحافظه

– model of Gilbert	- الگوی ژیلبرت
– two - way	- دوطرفه
– with memory	- باحافظه
channel coding	کدگذاری کانال
channel coding theroem	نظریه کدگذاری کانال
channel decoding	کدگشایی کانال
channel matrix	ماتریس کانال
Chaundy	چاندی
chosen - plaintext attack	اقدام با متن ساده -انتخابی
cipher	رمزی
– Caesar	- سزار
– substitution	- جانشین
– transposition	- انتقال
cipher system	سیستم رمزی
ciphertext	متن رمزی
ciphertext - only attack	اقدام تنها با متن رمزی
code	کد
– arithmetic	- حسابی
– block	- بلوکی
– Gilbert - Moore	- ژیلبرت- مور
– Hamming distance	- فاصلهٔ هامینگ
– Huffmann	- هافمن
– instantaneous	- لحظه‌ای (فوری)
– Morse	- مُرس
– non - singular	- ناویژه
– Shannon	- شانون
– Hamming	- هامینگ
– repetition	- تکرار
code efficiency	کارایی کد (کدکارا)

code rate	نرخ کد
code word	کدواژه
code word alphbet	الفبای کدواژه
coding	کدگذاری
– alphabetical extension	- توسعهٔ الفبایی
– source	- منبع
coding channel	کانال کدگذاری
coding strategies	استراتژی کدگذاری
coding theorem, Shannon' s	قضیهٔ کدگذاری، شانون
comma code	کد کاما
communication channel, multi - access	کانال ارتباطی، چند-مدخلی
communication model	الگوی ارتباطی
conditional amount of information	مقدار شرطی اطلاع
conditional information measure	اندازهٔ اطلاع شرطی
conditional probability	احتمال شرطی
conditional probability densities	چگالی احتمال شرطی
countinuous communication channel	کانال ارتباطی پیوسته
countinuous information measure	اندازهٔ اطلاع پیوسته
converse source coding theorem	وارون قضیهٔ کدگذاری منبع
correlation	همبستگی
correlation coefficient	ضریب همبستگی
covariance	کوواریانس
cryptanalysis	تحلیل رمزی
cryptography	رمزنگاری
cryptology	رمزشناسی
Csiszár	سیساب
cumulative distribution	توزیع تجمّعی
D - admissibility	مجاز بودن-D

Daroczy	داروچی
data compression	تراکم داده‌ها
data processing theorem	قضیهٔ پردازش داده‌ها
data reconstruction	بازسازی داده‌ها
data reduction	کاهش داده‌ها
decipherment	رمزگشایی
decoding channel	کانال کدگشایی
decryption	رمزگشایی
degraded broadcast channel	کانال پخش تضعیف شده
dependence redundancy	حشو وابسته
discrete channel	کانال گسسته
discrete communication channel	کانال ارتباطی گسسته
discrete information source	منبع اطلاع گسسته
– with memory	- باحافظه
discrete memoryless (information) source	منبع (اطلاع) بی‌حافظه گسسته
distortion, average	دگرشکلی، متوسط
distortion matrix	ماتریس دگرشکلی
distortion measure	اندازهٔ دگرشکلی
distribution, normal	توزیع، نرمال
effective information	اطلاع کارا
efficiency of code	کارایی کد
efficient	کارا
enciphering	رمزی کردن
– block	- بلوکی
– stream	- گروهی
encipherment	رمزگذاری
encryption	رمزی
ensemble	دسته

ergodic Markov chain	زنجیر مارکوف ارگودیک
error probability	احتمال خطا
error - correcting codes	کدهای تصحیح-خطا
errors - and - erasure channel	کانال خطا-و-پاک‌شدگی
expectation	امید (میانگین)
Fano code	کد فانو
Fano's inequality	نابرابری فانو
first coding theorem, Shannon's	قضیهٔ اوّل کدگذاری، شانون
frequency - divisison multi - access	تقسیم فرکانس چند-مدخلی
Gallager	گالاگر
gaussian broadcast channel	کانال پخش گاوسی
gaussian channel with memory	کانال گاوسی باحافظه
gaussian distribution	توزیع گاوسی
gaussian signal	سیگنال گاوسی
gaussian white noise	نوفهٔ سفید گاوسی
generator matrix	ماتریس تولید کننده
Gilbert - Moore code	کد ژیلبرت - مور
Hagelberger code	کد هاگلبرگر
Hamming codes	کدهای هامینگ
Hamming distance of a code	فاصله هامینگ یک کد
Hartley	هارتلی
Hartley's measure	اندازه هارتلی
High Definition TV	وضوح عالی تلویزیون
Hintikka	هینتیکا
homogeneous Markov chain	زنجیر مارکوف همگن
Huffman code	کد هافمن

inequality of Fano	نابرابری فانو
information	اطلاع
– effective	- مؤثر
– selective	- انتخابی
– syntactic	- ترکیبی
information measure	اندازه اطلاع
– axiomatic foundation	- اصول موضوعه
– conditional	- شرطی
– continuous	- پیوسته
– joint	- توأم
– marginal	- حاشیه‌ای
– mutual	- متقابل
– of Shamnon	- شانون
information power	توان اطلاع
information pragmatic	اطلاع عملی
information semantic	اطلاع مفهومی (اطلاع معانی)
information source, memoryless	منبع اطلاع، بدون حافظه
information transmission	ارسال اطلاع
information transmission theorem	قضیهٔ ارسال اطلاع
instantaneous code	کد لحظه‌ای (فوری)
Jacobian	ژاکوبی
joint amount of information	مقدار توأم اطلاع
joint cumulative distribution	توزیع تجمعی توأم
joint information measure	اندازه اطلاع توأم
joint probability	احتمال توأم (مشترک)
joint probability density	چگالی احتمال توأم
key	کلید

key appearance equivocation	کلید حضور ایهام
key equivocation	کلید ایهام
known - plaintext attack	اقدام با متن ساده-معلوم
Kraft's inequality	نابرابری کرافت
Lagrange's method	روش لاگرانژ
linear block code	کد بلوکی خطی
– binary	- دوتایی
MacKay	مک‌کی
marginal probability	احتمال حاشیه‌ای
Markov	مارکوف
Markov chain	زنجیر مارکوف
– amount of information	- اندازه اطلاع
Markov processes	فرایند مارکوف
maximum amount of information	مقدار ماکسیمم اطلاع
McLeod	ماک‌لود
mean	میانگین
measure of information	اندازهٔ اطلاع
memoryless channel	کانال بی‌حافظه
memoryless source	منبع بی‌حافظه
message	پیام
– most probable	- محتملترین
message equivocation	ایهام پیام
minimum amount of information	مقدار می‌نیمم اطلاع
model of Gilbert	الگوی ژیلبرت
modulation	مدولاسیون
Morse code	کد مُرس
most probable messages	محتملترین پیام

multi - access communication channel	کانال ارتباطی، چند-مدخلی
multi - access communication network	شبکهٔ ارتباطی، چند-مدخلی
multiterminal communication networks	شبکهٔ ارتباطی چندپایانه‌ای
mutual information measure	اندازهٔ اطلاع متقابل
N - dimensional gaussian distribution	توزیع گاوسی N-بعدی
nat	نَت
network, multiterminal	شبکهٔ، چندپایانه‌ای
network information theory	نظریهٔ اطلاع شبکه‌ای
noise	نوفه
noisy channels	کانالهای نوفه‌دار
non - gaussian white noise	نوفهٔ سفید غیرگاوسی
non - singular code	کد ناویژه
normal distribution	توزیع نرمال
number of most probable messages	تعداد محتملترین پیام
Nyquist	نای‌کویست
parity check matrix	ماتریس بررسی توازن
period length	طول دوره
plaintext	متن‌ساده
polyalphabetic substitution	جایگزینی چندالفبایی
power density spectrum	توان چگالی طیفی
pragmatic information	اطلاع عملی
prefix code	کد پیش‌وند
probabilistic experiment	آزمایش احتمالی
probability	احتمال
– conditional	- شرطی
– joint	- توأم
– marginal	- حاشیه‌ای

probability density functions	تابع چگالی احتمال
probability distribution	توزیع احتمال
probability theory	نظریهٔ احتمال
production	محصول
properties of the R(D) function	ویژگیهای تابع R(D)
rate distortion function	تابع نرخ دگرشکلی
– continuous	- پیوسته
– properties	- ویژگیهای
rate distortion theory	نظریهٔ نرخ دگرشکلی
rate of transmission	نرخ ارسال
realization	مسیر
redundancy	حشو
Renyi	رنی
repetition codes	کدهای تکراری
requirements on information measures	شرایط بر اندازه‌های اطلاع
sampling theorem	قضیهٔ نمونه‌گیری
security	اطمینان
selective information	اطلاع انتخابی
semantic information	اطلاع مفهومی (معانی)
Shannon	شانون
Shannon code	کد شانون
Shannon's coding theorem	قضیهٔ کدگذاری شانون
Shannon's first coding theorem	قضیهٔ اول کدگذاری شانون
Shannon's information measure	اندازهٔ اطلاع شانون
Shannon's second coding theorem	قضیهٔ دوم کدگذاری شانون
Shannon – McMillan theorem	قضیهٔ شانون مک‌میلان
sinc – function	تابع-سینوس‌وار
source alphabet	الفبای منبع

source coding	منبع کدگذاری
source coding theorem	قضیهٔ کدگذاری منبع
standard deviation	انحراف معیار
state diagram	نمودار حالت
state of a Markov chain	حالت زنجیر مارکوف
stationary	مانا (ایستا)
stationary signals	سیگنالهای مانا
stationary transition probabilities	احتمالهای تغییر حالت مانا
statistically independent	استقلال آماری
stochastic signals	سیگنالهای تصادفی
stream enciphering	رمزی کردن گروهی
strictly stationary signal	سیگنال اکیداً مانا
substitution cipher	رمز جایگزین
substitution cipher systems	سیستمهای رمز جایگزین
symbol distortion measure	اندازهٔ دگرشکلی نماد
syndrome	عارضه
syndrome coding	کدگذاری عارضه
syntactic information	اطلاع ترکیبی
theorem of Bayes	قضیهٔ بیز
time – division multi – access	تقسیم زمان-چند-مدخلی
Toeplitz matrix	ماتریس-تئوپولی
total redundancy	حشو کل
transition	تغییر حالت
transposition cipher	رمز انتقالی
transposition cipher systems	سیستمهای رمز انتقالی
trellis diagrams	دیاگرامهای داربستی
two – way channels	کانال دوطرفه

فاصلهٔ یکتایی	unicity distance
توزیع یکنواخت	uniform distribution
چگالی احتمال یکنواخت	uniform probability density
قابل رمزگشایی به طور یکتا	uniquely decodable
وان-در-لوب	Van der Lubbe
واریانس	variance
دیاگرام ون	Venn diagram
جدول ویگنر	Vigenere – tableau
سیگنال مانای ضعیف	weak stationary signal
وزن	weight
نوفهٔ سفید	white noise
- گاوسی	– gaussian
- غیرگاوسی	– non – gaussian
کانال- Z	Z – channel

راهنمای موضوعی

FERDOWSI UNIVERSITY OF MASHHAD

Publication No. 308

INFORMATION THEORY

Jan C. A. Van Der Lubbe

Translated by

DR. H.A. Azarnoosh

FERDOWSI UNIVERSITY PRESS

2001