مقاله کلاسیک

نظریهٔ ریاضی مخابرات

کلود شانُن

ترجمهٔ هرمز شهرزاد

مقالهٔ کلاسیک این شماره که عنوان اصلی آن
The mathematical theory of communication
است معمولاً سرآغاز بلامنازع بررسی ریاضی مفهوم «اطلاع» محسوب می‌شود. این مقاله اولین بار در سال ۱۹۴۸ در مجموعهٔ گزارشهای فنی آزمایشگاه بِل منتشر شد و پس از آن در مجموعه‌های گوناگونی تجدید چاپ شده است. کلود شانُن (Claude E. Shannon) نویسندهٔ مقاله در سال ۱۹۱۶ در ایالت میشیگان آمریکا متولد شد؛ در سال ۱۹۳۶ لیسانس مهندسی برق و در سال ۱۹۴۰ دکتری ریاضیات دریافت کرد. پس از گذراندن یک دورهٔ تحقیقاتی در مؤسسهٔ مطالعات عالی پرینستن به آزمایشگاه بِل پیوست و از سال ۱۹۵۷ به سمت استادی مهندسی برق ام. آی. تی. منصوب گردید. شانُن که اکنون بازنشسته است آثار مهمی در نظریه‌های اطلاعات و مخابرات، علوم کامپیوتر، نظریهٔ رمزنگاری، و طراحی مدار دارد.

اخیراً با ابداع روشهای متعددی در مدوله‌سازی[1] مثل PCM[2] و PPM[3] که پهنای باند را برای نسبت سیگنال به نویز تغییر می‌دهند، توجه به نظریهٔ عمومی مخابرات تشدید شده است. اساس چنین نظریه‌ای در مقالات مهم نایکویست [۱] و هارتلی [۲] در همین مقوله ذکر شده است. در مقالهٔ حاضر ما این نظریه را بسط می‌دهیم تا عوامل جدیدی علی‌الخصوص تأثیر نویز در کانال، امکان صرفه‌جویی بر مبنای ساختار آماری پیام اصلی، و طبیعت مقصد نهایی اطلاعات را در بر بگیرد.

مسألهٔ اساسی مخابرات این است که پیام منتخب در یک نقطه باید به‌طور دقیق یا تخمینی در نقطه‌ای دیگر بازسازی شود. غالبا پیامها دارای *معنا* هستند؛ یعنی در بارهٔ سیستمی با هویاتِ فیزیکی یا ذهنیِ مشخص‌اند یا طبق این سیستم با هم مرتبط‌اند. این جنبه‌های معنایی مخابرات به مسائل مهندسی ربطی ندارند. در اینجا جنبهٔ مهم آن است که پیام واقعی دقیقاً همان باشد که از یک مجموعهٔ پیامهای ممکن *انتخاب شده است*. این سیستم باید طوری طراحی شود که برای تک تک انتخابهای ممکن کارایی داشته باشد نه فقط برای آن که عملاً انتخاب خواهد شد، چرا که در زمان طراحیْ این موضوع نامعلوم است.

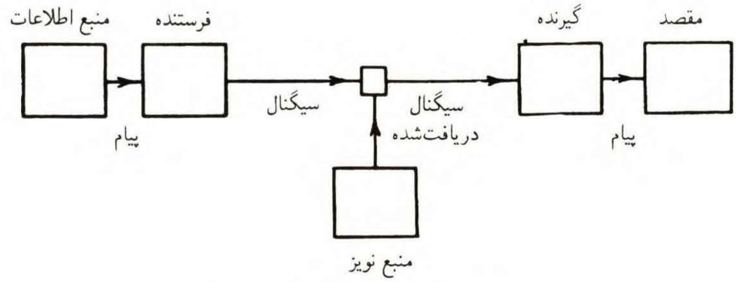
اگر تعداد پیامهای مجموعهٔ مذکور متناهی باشد، این تعداد یا هر تابع یکنوایی از آن را می‌توان به‌عنوان شاخص اطلاع حاصل از انتخاب یک پیام از این مجموعه تلقی کرد که برای تمام انتخابهای مشابه برابر است. همان‌گونه که هارتلی اشاره کرده، طبیعیترین انتخاب همان تابع لگاریتمی است. اگرچه زمانی که تأثیر آماری پیام را در نظر می‌گیریم و یا یک حوزهٔ پیوسته از پیامها داریم، این تعریف را باید تا حد زیادی تعمیم داد، معهذا ما در تمام موارد یک شاخص اساساً لگاریتمی را به‌کار خواهیم گرفت.

شاخص لگاریتمی به دلایل متعددی مناسبتر است:

۱. عملاً بیشتر به‌کار می‌آید. پارامترهای مهم مهندسی چون زمان، پهنای باند، تعداد رله‌ها و غیره غالباً متناسب با لگاریتم تعداد حالات ممکن تغییر می‌کنند. به‌عنوان مثال افزودن یک رله به یک گروه، تعداد حالات ممکن رله‌ها را دو برابر می‌کند. این کار در واقع یک واحد به لگاریتم پایهٔ ۲ی این تعداد اضافه می‌کند. دو برابر کردن زمانْ تعداد پیامهای ممکن را کم و بیش مجذور می‌کند و یا لگاریتم آن را دو برابر می‌کند و غیره.

۲. به‌عنوان یک شاخص مناسب به شهود ما نزدیکتر است. و این در ارتباط نزدیک با مورد (۱) است زیرا ما به‌طور شهودی چیزها را از طریق مقایسهٔ خطی با استانداردهای معمول اندازه می‌گیریم. به‌عنوان مثال هرکسی می‌فهمد که دو تا کارت سوراخ شده باید ظرفیت اطلاعات ذخیره شده در یک کارت را دو برابر کند و دو کانال مشابه، ظرفیت اطلاعاتِ انتقال یافته از یک کانال را دو برابر می‌کند.

1. modulation
2. Pulse-Code Modulation
3. Pulse-Position Modulation



شکل۱ نمودار شماتیک یک سیستم مخابراتی کلی.

۳. از نظر کاربرد ریاضی مناسبتر است. بسیاری از عملیاتِ حدی برحسب لگاریتم ساده هستند، اما برحسب تعداد حالات ممکن مستلزم تفصیلات دست و پاگیرتری می‌باشند.

انتخاب پایهٔ لگاریتم به منزلهٔ انتخاب واحد برای اندازه‌گیری اطلاعات است. اگر پایهٔ ۲ مورد استفاده قرار گیرد واحد نتایج را می‌توان ارقام دودویی یا به‌طور خلاصه‌تر، بیت[1] نامید که واژهٔ پیشنهادی توکی[2] است. وسیلهٔ با دو حالتِ ثابت مثل رله یا یک مدار فلیپ فلاپ می‌تواند یک بیت اطلاعات را ذخیره کند. پس N تا از این وسایل می‌توانند N بیت را ذخیره کنند، زیرا تعداد کل حالات ممکن ۲^N است و $\log_۲ ۲^N = N$. اگر از پایهٔ ۱۰ استفاده کنیم واحدها را می‌توان ارقام دهدهی نامید. از آنجا که

$$\log_۲ M = \log_{۱۰} M / \log_{۱۰} ۲$$
$$= ۳٫۳۲ \log_{۱۰} M$$

هر رقم دهدهی حدودا ۳۱/۳ بیت است. در ماشینهای حساب رومیزی هر چرخِ رقمی دارای ده حالت ثابت است و بنابراین دارای ظرفیت ذخیرهٔ یک رقم دهدهی است. در زمینهٔ آنالیز، جایی که مشتق و انتگرال دخیل هستند، پایهٔ e گاهی اوقات مفید است. واحدهای به‌دست آمدهٔ اطلاعات را واحدهای طبیعی می‌نامند. تغییر از پایهٔ a به پایهٔ b تنها مستلزم عمل ضرب در $\log_b a$ است.

منظور از سیستم مخابراتی، آن نوع سیستمی است که به‌صورت شماتیک در شکل ۱ نشان داده شده و اساسا شامل پنج قسمت است:

۱. یک منبع *اطلاعات* که تولیدکنندهٔ یک پیام یا دنباله‌ای از پیامهاست که قراراست به پایانهٔ[3] مقصد ارسال شود. پیام ممکن است انواع مختلف داشته باشد: (الف) دنباله‌ای از حروف مثل تلگراف یا سیستم تله تایپ؛ (ب) یک تابع یک متغیره از زمان، $f(t)$، مثل آنچه در رادیو یا تلفن می‌بینیم؛ (پ) تابعی از زمان و سایر متغیرها مثل آنچه در تلویزیون سیاه و سفید وجود دارد ــ در اینجا پیام را می‌توان به‌عنوان تابع $f(x,y,t)$ از دو مؤلفهٔ فضا و یک مؤلفهٔ زمان فرض کرد: شدت نور در زمان t در نقطهٔ (x,y) در لامپ تصویر؛ (ت) دو تابع زمانی یا بیشتر مثل $f(t)$ و $g(t)$ و $h(t)$، و این در مورد انتقال صدا به شکل «سه‌بعدی» است و یا وقتی که سیستم متضمن ارائهٔ کانالهای متعدد مجزا به‌صورت مرکب[4] است؛ (ث) توابع متعدد از چند متغیر، مثلا در

1. bit 2. J. W. Tukey 3. terminal 4. multiplex

تلویزیون رنگی، پیام شامل سه تابع $f(x,y,t)$ و $g(x,y,t)$ و $h(x,y,t)$ است که در پیوستار سه‌بعدی تعریف می‌شوند ــ و البته می‌توانیم از این سه تابع به‌عنوان مؤلفه‌های یک میدان برداری که در آن محدوده تعریف شده است یاد کنیم ــ به‌طور مشابه، بسیاری از فرستنده‌های تلویزیون‌های سیاه و سفید هم «پیامهایی» تولید می‌کنند که از توابع سه‌متغیره تشکیل شده است؛ (ج) ترکیبات مختلفی از موارد فوق هم روی می‌دهد، به‌عنوان مثال تلویزیونی که با کانال صدا همراه است.

۲. یک فرستنده که به‌نحوی روی پیام کار می‌کند که سیگنال مناسب برای عبور از کانالِ جریان را تولید کند. در ارتباط تلفنی این کار تنها شامل تبدیل فشار ارتعاشات صدا به الکتریسیتهٔ متناسب است. در تلگراف از عمل کدگذاری[1] استفاده می‌کنیم که در قبال پیام یک سری خط و نقطه و فاصله بر روی کانال می‌فرستد. در یک سیستم مرکب PCM، توابع متفاوت گفتاری باید نمونه‌برداری، فشرده، گسسته[2] و کدگذاری شوند و بالاخره برای تولید سیگنال، به شکل مناسب جایگذاری گردند. سیستمهای کُدگذار صوت[3]، تلویزیون و مدوله‌سازی فرکانسی نمونه‌های دیگری از عملیات پیچیده‌ای هستند که برای دستیابی به سیگنال، بر پیام اعمال می‌شوند.

۳. *کانال* صرفاً واسطه‌ای است که برای انتقال سیگنال از فرستنده به گیرنده مورد استفاده قرار می‌گیرد. این کانال ممکن است یک جفت سیم، یک کابل هم‌محور[4]، یک باند فرکانس رادیویی، یک باریکهٔ نور و غیره باشد. در طی انتقال، یا در یکی از پایانه‌ها، ممکن است سیگنال به‌واسطهٔ نویز مخدوش گردد. این موضوع به‌صورت شماتیک در شکل ۱ نشان داده شده است که در آن منبع نویز روی سیگنال انتقالی اثر می‌گذارد تا سیگنال دریافتی تولید شود.

۴. *گیرنده* معمولاً عکس کاری را که فرستنده انجام می‌دهد، اجرا می‌کند، یعنی پیام را از روی سیگنال بازسازی می‌کند.

۵. مقصد شخصی (یا چیزی) است که قرار است پیام به او برسد.

قصد ما این است که بعضی مسائل کلی را که در سیستمهای مخابراتی وجود دارد بررسی کنیم. برای انجام این کار، ابتدا لازم است که عناصر متعدد درگیر در مسأله را به‌عنوان اشیایی ریاضی ارائه کنیم که به شکل مناسبی از مصادیق فیزیکی خودشان منتزع شده باشند. سیستمهای مخابراتی را تقریبا می‌توان به سه نوع اصلی تقسیم نمود: گسسته، پیوسته و مختلط. منظور ما

1. encoding 2. quantized 3. vocoder 4. coaxial cable

از سیستم گسسته این است که هم پیام و هم سیگنال هر دو دنباله‌ای از نمادهای گسسته هستند. مثال بارز این نوعْ تلگراف است که در آن پیام از یک دنبالهٔ حروف و سیگنال از یک دنبالهٔ نقطه، خط و فاصله تشکیل شده است. سیستم پیوسته سیستمی است که در آن پیام وسیگنال هردو به‌عنوان توابعی پیوسته رفتار می‌کنند، مثل رادیو یا تلویزیون. سیستم مختلط، سیستمی است که در آن متغیرهای گسسته و پیوسته هر دو ظاهر می‌گردند، مثل انتقال گفتار با روش PCM.

در ابتدا مورد گسسته را بررسی می‌کنیم. این مورد نه تنها در نظریهٔ مخابرات بلکه در نظریهٔ ماشینهای محاسب، طراحی تبادلات تلفنی و زمینه‌های دیگر نیز کاربرد دارد. به‌علاوه، مورد گسسته پایه‌ای است برای موارد پیوسته و مختلط که در دومین بخش همین مقاله به آنها می‌پردازیم.

سیستمهای گسستهٔ بدون نویز

۱. کانال گسستهٔ بدون نویز

تله‌تایپ و تلگراف دو نمونهٔ ساده از کانال گسسته برای انتقال اطلاعات هستند. کانال گسسته عموماً به مفهوم سیستمی است که از طریق آن یک سری نمادهای ابتدایی انتخاب شده از یک مجموعهٔ متناهی S_1، ...، S_n را می‌توان از یک نقطه به نقطه‌ای دیگر انتقال داد. زمان مشخص ارسال هر یک از S_iها t_i ثانیه است (که الزاماً برای S_iهای متفاوت یکی نیست، برای مثال نقطه‌ها و خطها در تلگراف). و البته لزومی هم ندارد که همهٔ دنباله‌های ممکن S_iها قابل انتقال در سیستم باشند؛ بلکه ممکن است تنها دنباله‌های خاصی مجاز شمرده شوند. اینها سیگنالهای ممکن برای کانال هستند. مثلاً فرض کنید که در تلگراف نمادها عبارت باشند از: (۱) یک نقطه، که عبارت است از بسته شدن و باز شدن مدار، هریک به میزان یک واحد زمانی؛ (۲) یک خط، که از سه واحد زمانی مدار بسته و یک واحد باز تشکیل شده است؛ (۳) یک فاصلهٔ حرفی که مثلاً سه واحد زمانی مدار باز است؛ (۴) فاصلهٔ بین کلمات که شش واحد زمانی مدار باز است. ما می‌توانیم این محدودیت را بر روی دنباله‌های مجاز اعمال کنیم که فاصله‌ها به‌دنبال یکدیگر قرار نگیرند (زیرا اگر دو فاصلهٔ بین حروف در کنار هم قرار بگیرند با یک فاصله بین کلمات برابر خواهند بود). سؤالی که ما اینک به آن می‌پردازیم این است که چگونه می‌توان ظرفیت انتقال اطلاعات را برای چنین کانالی اندازه‌گیری کرد.

در مورد تله‌تایپ که کلیهٔ نمادها مدت ارسال مشابه دارند و کلیهٔ دنباله‌های هر یک از ۳۲ نماد موجود، مجاز شمرده می‌شوند جواب ساده است. هر نماد نمایانگر پنج بیت اطلاعات است. اگر سیستم n نماد در ثانیه ارسال کند طبیعی است که بگوییم ظرفیت کانال $5n$ بیت در ثانیه است. این حرف به این معنا نیست که کانال تله‌تایپ همواره اطلاعات را به این میزان ارسال می‌کند، بلکه این حداکثر سرعت ممکن است و همان‌گونه که بعداً خواهیم دید بسته به منبع اطلاعاتی که کانال از آن تغذیه می‌شود، میزان واقعی ممکن است به این حداکثر برسد یا نرسد.

در حالت کلیتر که نمادها غیر هم‌طول‌اند و محدودیتهایی بر روی دنباله‌های مجاز وجود دارد، تعریف زیر را ارائه می‌کنیم: C، ظرفیت یک کانال گسسته، از رابطهٔ زیر به‌دست می‌آید

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T}$$

که در آن $N(T)$ تعداد سیگنالهای مُجازی است که طول مدتی برابر T دارند.

به سادگی دیده می‌شود که در مورد تله‌تایپ این عبارت به شکل نتیجهٔ قبلی در می‌آید. می‌توان نشان داد که حد مذکور در اغلبِ حالتهای موردِ توجهٔ موجود و یک عدد متناهی است. حال فرض کنید که کلیهٔ دنباله‌های نمادهای S_1، ...، S_n مجاز هستند و به‌ترتیب دارای مدت زمان ارسال t_1، ...، t_n می‌باشند. ظرفیت کانال چقدر است؟ اگر $N(t)$ بیانگر تعداد دنباله‌هایی باشد که مدت ارسالشان برابر t است، داریم

$$N(t) = N(t-t_1) + N(t-t_2) + \ldots + N(t-t_n)$$

این مجموع برابر است با مجموع تعداد دنباله‌هایی که به هر یک از نمادهای S_1، S_2، ...، S_n ختم می‌شوند که اینها به‌ترتیب همان

$$N(t-t_1), N(t-t_2), \ldots, N(t-t_n)$$

هستند. بر اساس حکم معروفی در حساب تفاضلات متناهی، $N(t)$ برای tهای بزرگ دارای مجانبی به‌صورت AX_0^t می‌باشد که در آن A یک مقدار ثابت و X_0 بزرگترین ریشهٔ حقیقیِ معادلهٔ مشخصهٔ ذیل است

$$X^{-t_1} + X^{-t_2} + \ldots + X^{-t_n} = 1$$

و بنابراین

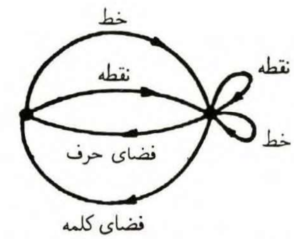
$$C = \lim_{T\to\infty} \frac{\log AX_0^T}{T} = \log X_0$$

در حالتی نیز که دنباله‌های مجاز محدودیتهایی دارند کماکان می‌توانیم یک معادلهٔ تفاضلی از این دست تشکیل داده C را از روی معادلهٔ مشخصهٔ آن پیدا کنیم. مثلا در مورد تلگراف که پیشتر به آن اشاره شد داریم

$$\begin{aligned} N(t) &= N(t-2) + N(t-4) + N(t-5) \\ &\quad + N(t-7) + N(t-8) + N(t-10) \end{aligned}$$

که همان‌طور که می‌بینیم با منظور کردن دنباله‌های نمادها بر اساس نماد آخر یا ماقبل آخر به‌دست می‌آید. بنابراین وقتی که μ_0 ریشهٔ مثبت معادلهٔ $1 = \mu^2 + \mu^4 + \mu^5 + \mu^7 + \mu^8 + \mu^{10}$ باشد C برابر با $-\log \mu_0$ خواهد بود. از حل این معادله مقدار $C = 0.539$ به‌دست می‌آید.

یکی از عمومی‌ترین انواع محدودیتهایی که می‌توان بر روی دنباله‌های مجاز اعمال نمود به قرار زیر است: تعدادی از حالتهای ممکن a_1، a_2، ...، a_m را اختیار می‌کنیم. برای هر حالت فقط نمادهای مشخصی از مجموعهٔ S_1، S_2، ...، S_n می‌توانند منتقل گردند (زیر مجموعه‌های متفاوت برای وضعیتهای مختلف). هنگام ارسال هریک از این نمادها، بسته به حالت قبلی و نماد مشخص ارسالی، حالت قبلی تبدیل به حالت جدید می‌گردد.



شکل۲ نمایش ترسیمی محدودیتهای روی نمادهای تلگراف.

تلگراف یک مثال ساده از این امر است. بسته به اینکه آخرین نماد ارسالی فاصله باشد یا خیر، دو حالت موجود است. اگر چنین باشد، تنها نمادهایی که می‌توانند متعاقباً ارسال شوند نقطه و خط هستند و حالت همواره تغییر می‌کند. اگر چنین نباشد، هر یک از نمادها می‌توانند ارسال گردند و حالت فقط در صورتی تغییر می‌کند که یک فاصله ارسال شود و درغیر این صورت بدون تغییر باقی می‌ماند. این شرایط را می‌توان در یک گراف خطی مانند شکل ۲ نشان داد. رئوس متناظرند با حالتها، و خطوط نشانگر نمادهای ممکن در یک حالت و حالتِ حاصل هستند. در پیوست ۱ نشان داده شده است که چنانچه شرایطِ اعمالی بر روی دنباله‌ها را بتوان به این صورت توصیف کرد، C موجود و طبق حکمِ زیر قابل محاسبه است:

قضیهٔ ۱. *اگر $b_{ij}^{(s)}$ مدت sامین نماد مجاز در حالت i باشد که ما را به حالت j می‌برد، آنگاه C، ظرفیت کانال، برابر است با $\log(W)$، که W بزرگترین ریشهٔ حقیقیِ معادله دترمینانی زیر است:*

$$\left|\sum_s W^{-b_{ij}^{(s)}} - \delta_{ij}\right| = 0$$

که در آن $\delta_{ij} = 1$ اگر $i = j$، و در غیر این‌صورت δ صفر خواهد بود.

برای مثال، در مورد تلگراف (شکل ۲) این دترمینان به‌صورت زیر است

$$\begin{vmatrix} -1 & (W^{-2} + W^{-4}) \\ (W^{-3} + W^{-6}) & (W^{-2} + W^{-4} - 1) \end{vmatrix} = 0$$

که بسط این عبارت ما را به همان معادله‌ای که قبلاً برای این شرایط ذکر شد، می‌رساند.

۲. منبع گسستهٔ اطلاعات

مشاهده کردیم که تحت شرایطی بسیار کلی، لگاریتم تعداد سیگنالهای ممکن در کانالِ گسسته به‌طور خطی با زمان افزایش می‌یابد. ظرفیت انتقال اطلاعات را هم می‌توان با ارائهٔ همین آهنگِ افزایش مشخص کرد، که همانا تعداد بیتهای لازم در ثانیه برای مشخص کردن سیگنال خاصِ مورد استفاده است.

حالا به منبع اطلاعات می‌پردازیم. چطور می‌توان منبع اطلاعات را از دیدگاه ریاضی توصیف کرد، و در یک منبع مفروض چه مقدار اطلاعات برحسب بیت در ثانیه تولید می‌شود؟ نکتهٔ اصلی این بحثْ تأثیر معلومات آماری دربارهٔ منبع در استفاده از کدگذاری مناسبِ اطلاعات برای کم‌کردن ظرفیت لازم در کانال است. به‌عنوان مثال در تلگراف، پیامها از طریق دنباله‌هایی از حروف ارسال می‌شوند. این دنباله‌ها کاملاً هم تصادفی نیستند. در حالتِ کلی آنها جملاتی می‌سازند و مثلاً ساختار آماریِ زبان انگلیسی را دارا هستند. حرف E بیشتر از حرف Q به چشم می‌خورد، دنبالهٔ TH بیشتر از XP، و غیره. وجود چنین ساختاری اجازه می‌دهد که با کدگذاری مناسب دنبالهٔ پیامها در هنگام تبدیل به دنبالهٔسیگنالها، در زمان (یا در ظرفیت کانال) صرفه‌جویی کرد. این کار تاحدی در تلگراف با استفاده از کوتاهترین نماد کانال یعنی یک نقطه برای نمایش معمولترین حرف انگلیسی، E، صورت گرفته است؛ در حالی‌که حروف غیرمعمولتر Q، X، Z با استفاده از دنباله‌های طولانیتری از نقطه‌ها و خطها نمایش داده می‌شوند. به‌علاوه در کُدهای یک زمینهٔ مشخص بازرگانی، که کلمات و عبارات معمول با گروهِ کُدهای چهار یا پنج حرفی بیان می‌شوند و کاهش قابل ملاحظه‌ای را در متوسط زمان سبب می‌گردند، این معنا پیشتر می‌رود. در تلگرامهای استاندارد شدهٔ تبریک و سالگرد که امروزه به‌کار می‌روند این پیشروی تا جایی ادامه می‌یابد که یک یا دو جمله با تعداد دنباله‌های نسبتاً کوتاهی کُدگذاری شوند.

می‌توانیم فرض کنیم که منبعِ گسستهٔ پیام را به‌صورت نماد به نماد تولید می‌کند، به‌طوری که انتخاب نماد بعدی عموماً با احتمال مشخصی به نمادهای قبلی و خودِ نمادهای مورد بحث بستگی دارد. یک سیستم فیزیکی یا مدلی ریاضی از یک سیستم که چنین دنباله‌هایی از نمادها را تولید می‌کند که آن هم به‌وسیلهٔ مجموعه‌ای از احتمالات تعیین می‌گردد، به‌عنوان فرایندی تصادفی شناخته می‌شود [۳]. بنابراین می‌توانیم یک منبع گسسته را با فرایندی تصادفی نمایش دهیم. از طرف دیگر، هر فرایند تصادفی را که دنبالهٔ گسسته‌ای از نمادها را از یک مجموعهٔ نامتناهی برمی‌گزیند می‌توان به‌عنوان منبعی گسسته در نظر گرفت که موارد زیر را شامل خواهد شد:

۱. زبانهای طبیعی نوشتاری مثل انگلیسی، آلمانی، چینی.

۲. منابع پیوستهٔ اطلاعات که از طریق برخی فرایندهای گسسته‌سازی به شکلِ گسسته تحویل شده‌اند. برای مثال، گفتارِ گسسته‌شده در یک سیستم PCM، یا علائم تلویزیونیِ گسسته‌شده.

۳. موارد ریاضی که صرفاً معرف فرایند تصادفی مجردی هستند که مولد دنباله‌ای از نمادهاست. مواردی که ذیلاً به‌آنها اشاره می‌شود از نوع اخیرند.

(الف) فرض کنید پنج حرف A و B و C و D و E را داریم که هر کدام به‌احتمال ۲ر۰ انتخاب می‌شوند و انتخابهای بعدی هم مستقل هستند. این ما را به دنباله‌ای که مثالی نوعی از آن آمده است، هدایت می‌کند.

B D C B C E C C C A D C B D D A A E C E E A
A B B D A E E C A C E E B A E E C B C E A D

این مثال با استفاده از جدول اعداد تصادفی تولید شده است [۴].

(ب) با استفاده از همان پنج حرف در نظر بگیرید که احتمالات به‌ترتیب ۰٫۴ و ۰٫۱ و ۰٫۲ و ۰٫۲ و ۰٫۱ باشند و انتخابهای بعدی هم مستقلاً انجام گیرد. یک پیام نوعی از این منبع بدین صورت خواهد بود

A A A C D C B D C E A A D A D A C E D A
E A D C A B E D A D D C E C A A A A A D

(پ) هنگامی که انتخابهای بعدی مستقلاً انجام نگیرد و احتمالهای آنها به حروف قبلی بستگی داشته باشد ساختار پیچیده‌تری به‌دست می‌آید. در ساده‌ترین مورد از این دست، یک انتخاب فقط به حرف قبلی و نه حروف ماقبل آن بستگی دارد. چنین ساختار آماری را می‌توان به‌صورت مجموعه‌ای از احتمالات انتقال، $p_i(j)$، یعنی احتمال آمدن حرف j پس از حرف i، توصیف نمود به‌گونه‌ای که i و j روی محدودهٔ تمام نمادهای ممکن تغییر می‌کنند. دومین روشی که این ساختار را توصیف می‌کند به‌دست دادن احتمالات «زوج»[1] $p(i, j)$ یا همان بسامد نسبی زوج ij می‌باشد. بسامدهای حرفی $p(i)$ (احتمال حرف i)، احتمالات انتقال $p_i(j)$ و احتمالات زوج $p(i, j)$ با فرمولهای زیر با یکدیگر مرتبط می‌گردند

$$p(i) = \sum_j p(i, j) = \sum_j p(j, i) = \sum_j p(j) p_j(i)$$

$$p(i, j) = p(i) p_i(j)$$

$$\sum_j p_i(j) = \sum_i p(i) = \sum_{i,j} p(i, j) = ۱$$

به‌عنوان یک مثال مشخص فرض کنید که سه حرف A و B و C با جداول احتمالات زیر داده شده‌اند

$p_i(j)$		A	B	C
i	A	۰	$\frac{۴}{۵}$	$\frac{۱}{۵}$
	B	$\frac{۱}{۲}$	$\frac{۱}{۲}$	۰
	C	$\frac{۱}{۲}$	$\frac{۲}{۵}$	$\frac{۱}{۱۰}$

i	$p(i)$
A	$\frac{۹}{۲۷}$
B	$\frac{۱۶}{۲۷}$
C	$\frac{۲}{۲۷}$

$p(i, j)$		A	B	C
i	A	۰	$\frac{۴}{۱۵}$	$\frac{۱}{۱۵}$
	B	$\frac{۸}{۲۷}$	$\frac{۸}{۲۷}$	۰
	C	$\frac{۱}{۲۷}$	$\frac{۴}{۱۳۵}$	$\frac{۱}{۱۳۵}$

یک پیام نوعی از این منبع به‌صورت زیر است

A B B A B A B A B A B A B A B B B A B B.
B B B A B A B A B A B A B B B A C A C A
B B A B B B B A B B A B A C B B B A B A

1. digram

قدم بعدی در افزایش پیچیدگی، بسامدهای سه‌تایی و نه بیشتر را به میان می‌آورد، به‌طوری که انتخاب یک حرف به دو حرف قبلی و نه مابقیِ پیام قبل از آن نقطه بستگی خواهد داشت و در آن مجموعه‌ای از بسامدهای سه‌تایی[1] $p(i, j, k)$ یا معادلاً مجموعه‌ای از احتمالات انتقالِ $p_{ij}(k)$ مورد نیاز خواهد بود. با ادامهٔ این راه فرایندهای تصادفی به‌مراتب پیچیده‌تری را می‌توان به‌دست آورد. در حالت کلی nتایی[2]، مجموعه‌ای از احتمالات nتایی $p(i_1, i_2, \ldots, i_n)$ یا احتمالاتِ انتقال $p_{i_1, i_2, \ldots, i_{n-1}}(i_n)$ برای توصیف ساختار آماری سیستم مورد نیاز است.

(ت) فرایندهای تصادفی را می‌توان به‌عنوان تولیدکنندهٔ متنی که شامل دنبالهٔ «کلمات» است نیز تعریف نمود. فرض کنید در زبانی پنج حرف A، B، C، D، E و ۱۶ «کلمه» با احتمالات مربوط به آنها موجود است

۰٫۱۰	A	۰٫۱۶	BEBE	۰٫۱۱	CABED	۰٫۰۴	DEB
۰٫۰۴	ADEB	۰٫۰۴	BED	۰٫۰۵	CEED	۰٫۱۵	DEED
۰٫۰۵	ADEE	۰٫۰۲	BEED	۰٫۰۸	DAB	۰٫۰۱	EAB
۰٫۰۱	BADD	۰٫۰۵	CA	۰٫۰۴	DAD	۰٫۰۵	EE

همچنین فرض کنید که «کلمات» متوالی مستقلاً انتخاب شده و با یک فاصله از هم جدا می‌شوند. یک پیام نوعی می‌تواند چنین باشد:

DAB EE A BEBE DEED DEB ADEE ADEE
EE DEB BEBE BEBE BEBE ADEE BED
DEED DEED CEED ADEE A DEED DEED
BEBE CABED BEBE BED DAB DEED ADEB

چنانچه تمام کلمات دارای طول متناهی باشند این فرایند معادل یکی از انواع قبلی خواهد بود ولی تشریح آن برحسب ساختار و احتمالاتِ کلمه‌ای می‌تواند ساده‌تر باشد. ما همچنین می‌توانیم با مطرح کردن احتمالات انتقال بین کلمات و غیره، مسأله را تعمیم بدهیم.

این زبانهای مصنوعی در ساختن مسائل و مثالهای ساده‌ای که امکانات مختلف را نشان می‌دهند، کاربرد دارند. ما همچنین می‌توانیم به وسیلهٔ یک سری از زبانهای مصنوعیِ ساده یک زبان طبیعی را تقریب بزنیم. تقریبِ مرتبهٔ صفر با انتخاب مستقل و متساوی‌الاحتمالِ کلیهٔ حروف به‌دست می‌آید. تقریب مرتبهٔ اول با انتخاب مستقل حروف متوالی ولی دارای احتمال مساوی با احتمالِ هر حرف در زبان طبیعی به‌دست می‌آید [۵]. مثلاً در تقریب مرتبهٔ اول انگلیسی، حرف E با احتمال ۰٫۱۲ انتخاب می‌شود (بسامدش در انگلیسی معمولی) و حرف W با احتمال ۰٫۰۲، اما حروف همسایه هیچ تأثیری بر روی هم ندارند و هیچ تمایلی به ایجاد زوجهای ارجحی مانند TH، ED و غیره وجود ندارد. در تقریب مرتبهٔ دوم ساختار زوجی به میان می‌آید. بعد از انتخاب یک حرف، حرف بعدی مطابق با بسامدِ حروفِ مختلفی که به‌دنبال

1. trigram 2. n-gram

اولی می‌آیند، انتخاب می‌شود. این کار به یک جدول از بسامدهای زوجی $p_i(j)$ نیازمند است. در تقریب مرتبهٔ سوم، ساختار سه‌تایی مطرح می‌گردد. هر حرف با احتمالاتی که به دو حرف ماقبل آن بستگی دارد انتخاب می‌شود.

۳. زنجیرهٔ تقریبهای انگلیسی

برای اینکه تصوری از این مطلب داشته باشید که چگونه این زنجیرهٔ فرایندها به یک زبان نزدیک می‌شوند، دنباله‌هایی نوعی از تقریبهای زبان انگلیسی ایجاد و ذیلاً داده شده‌اند. در تمام موارد ما یک «الفبای» ۲۷ نمادی را که شامل ۲۶ حرف و یک فاصله است در نظر گرفته‌ایم:

۱. تقریب مرتبهٔ صفر (نمادهای مستقل و هم احتمال)

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ
FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD

۲. تقریب مرتبهٔ اول (نمادهای مستقل ولی با بسامدهای متن انگلیسی)

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA
TH EEI ALHENHTTPA OOBTTVA NAH BRL

۳. تقریب مرتبهٔ دوم (ساختار زوجی مانند زبان انگلیسی)

ON IE ANTSOUTINYS ARE T INCTORE ST BE S D-
EAMY ACHIN D ILONASIVE TUCOOWE AT TEAS-
ONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

۴. تقریب مرتبهٔ سوم (ساختار سه‌تایی مانند زبان انگلیسی)

IN NO IST LAT WHEY CRATICT FROURE
BIRS GROCID PONDENOME OF DEMONSTURES
OF THE REPTAGIN IS REGOACTIONA OF CRE

۵. تقریب کلمه‌ای مرتبهٔ اول. به‌جای ادامهٔ کار با ساختار چهارتایی، ... و nتایی، بهتر و ساده‌تر است که از این نقطه جهشی به‌سمت واحدهای کلمه‌ای بکنیم. دراینجا کلمات مستقلاً ولی با بسامدهای مناسب خودشان انتخاب می‌شوند.

REPRESENTING AND SPEEDILY IS AN GOOD
APT OR COME CAN DIFFERENT NATURAL
HERE HE THE A IN CAME THE TO
OF TO EXPERT GRAY COME TO FURNISHES
THE LINE MESSAGE HAD BE THESE

۶. تقریب کلمه‌ای مرتبهٔ دوم. احتمالات انتقال کلمات صحیح است ولی هیچ ساختار اضافی را شامل نمی‌گردد.

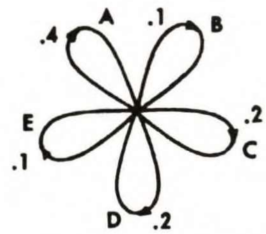
THE HEAD AND IN FRONTAL ATTACK ON AN E-
NGLISH WRITER THAT THE CHARACTER OF
THIS POINT IS THERE FORE ANOTHER
METHOD FOR THE LETTERS THAT THE TIME
OF WHO EVER TOLD THE PROBLEM FOR
AN UNEXPECTED

شباهت به متن انگلیسی معمولی به‌طور قابل ملاحظه‌ای در هر یک از مراحل بالا افزایش می‌یابد. توجه داشته باشید که این نمونه‌ها از ساختار بسیار خوبی برخوردار هستند، تقریباً دو برابر حدی که هنگام ایجادشان انتظار می‌رفت. مثلاً در مورد شمارهٔ (۳)، فرایند آماری. ارائهٔ متن قابل قبولی را برای دنباله‌های دو حرفی تضمین می‌کند، اما با این حال دنباله‌های چهارحرفی این نمونه را معمولاً می‌توان در جملاتِ خوب گنجانید. در مورد شمارهٔ (۶)، دنباله‌های چهار کلمه‌ای یا بیشتر را به‌آسانی می‌توان در جملات بدون ساختارهای غیرمعمول جایگزین کرد. دنبالهٔ خاص ده‌کلمه‌ایِ «attack on an English writer that the character of this» به هیچ عنوان غیر معقول نیست. پس معلوم می‌شود که یک فرایند تصادفیِ به نسبت پیچیده، بیان رضایت‌بخشی از یک منبع گسسته به‌دست خواهد داد.

دو نمونهٔ اول این بخش با استفاده از یک کتابِ اعداد تصادفی و بسامدهای حروف (برای نمونه ۲) ساخته شده است. همچنین این روش می‌تواند برای موارد (۳)، (۴) و (۵) ــ تا جایی‌که جداول بسامدهای زوجی و سه‌تایی و کلمه‌ای موجود باشند ــ ادامه پیدا کند، اما در عوض از روشی معادل ولی ساده‌تر استفاده شده است. مثلاً برای ساخت مثال (۳) تنها کافی است که به‌طور تصادفی کتابی را باز نمود و به‌طور تصادفی از همان صفحه حرفی را برگزید. این حرف ثبت می‌شود. پس از آن صفحهٔ دیگری از کتاب باز می‌گردد و خوانده می‌شود تا زمانی که به حرف مذکور برسیم. سپس حرف بعدی هم ثبت می‌شود. به صفحه‌ای دیگر از کتاب می‌رویم و این حرف دوم را جستجو کرده و حرف بعدی را ثبت می‌کنیم و به همین ترتیب. فرایند مشابهی برای موارد (۴)، (۵) و (۶) مورد استفاده قرار گرفته است. بنا کردن تقریبهای بعدی بسیار جالب خواهد بود، ولی زحمتِ این کار در مرحلهٔ بعدی فوق‌العاده افزایش می‌یابد.

۴. نمایش ترسیمیِ یک فرایند مارکُف

فرایندهای تصادفی از نوعی که در بالا تشریح شد در ریاضیات به‌عنوان فرایندهای گسستهٔ مارکُف شناخته می‌شوند و در نوشته‌های زیادی مورد بررسی قرار گرفته‌اند [۶]. حالت کلی آن می‌تواند به شکل زیر باشد: تعداد متناهی از «حالتها»ی سیستم: S_1، S_2، ...، S_n موجود است. به‌علاوه مجموعه‌ای از احتمالات انتقال، $p_i(j)$، (احتمال اینکه اگر سیستم در حالت S_i باشد، سپس به‌حالت S_j خواهد رفت) نیز موجود است. برای تبدیل این فرایندِ مارکُف به یک منبع اطلاعات، تنها لازم است فرض کنیم که برای انتقال



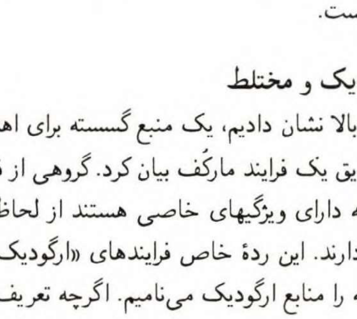
شکل۳ گراف مربوط به منبع مثال ب.

از یک حالت به حالتی دیگر، یک حرف تولید می‌گردد. حالتهای مذکور در تناظر با «ماندهٔ تأثیر» حروف قبلی خواهند بود.

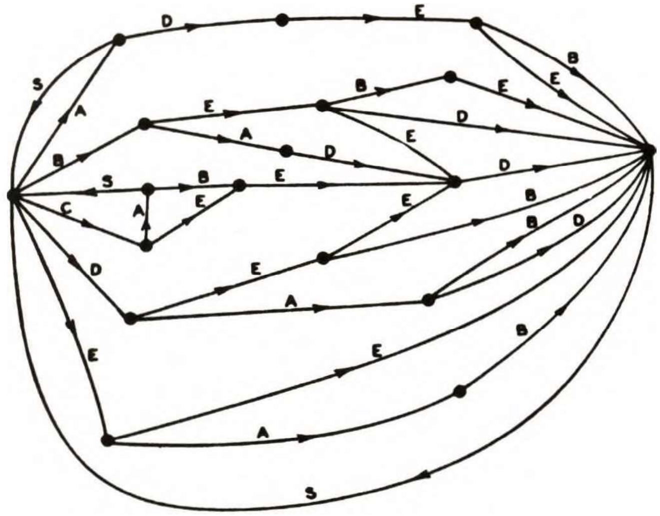
این وضعیت را می‌توان به شکل ترسیمی، همچنان که در اشکال ۳ ، ۴ و ۵ نشان داده شده است، نمایش داد. «حالتها» رئوس گراف هستند و احتمالات و حروف تولید شده برای یک انتقال، در کنار یال مربوطْ داده شده‌اند. شکل ۳ مربوط به مثالِ (ب)ی بخش ۲ و شکل ۴ مربوط به مثالِ (پ) است. در شکل ۳ از آنجایی که حروف بعدی مستقل‌اند فقط یک حالت موجود است. در شکل ۴ تعداد حالتها با تعداد حروف یکی است. چنانچه مثالی از سه‌تایی‌ها ساخته شده‌بود، متناظر با تعداد جفتهای حروف ممکن ماقبل آن حرفی که انتخاب می‌شد، حداکثر n^{2} حالت می‌داشت. شکل ۵ یک گراف برای مورد ساختار کلمه‌ای مثال (ت) است. که در آن S مربوط به نماد «فاصله» است.

۵. منابع ارگودیک و مختلط

همان‌گونه که در بالا نشان دادیم، یک منبع گسسته برای اهداف مورد نظر ما را می‌توان از طریق یک فرایند مارکُف بیان کرد. گروهی از فرایندهای گسستهٔ مارکُفِ ممکن که دارای ویژگیهای خاصی هستند از لحاظ نظریهٔ مخابرات اهمیت ویژه‌ای دارند. این ردهٔ خاص فرایندهای «ارگودیک» می‌باشند، و ما هم منابع مربوطه را منابع ارگودیک می‌نامیم. اگرچه تعریف دقیقِ فرایندهای ارگودیک تا حدی پیچیده است، معهذا ایدهٔ کلی ساده است. در یک فرایند ارگودیک هر دنبالهٔ تولید شده از این فرایندْ دارای ویژگیهای آماری یکسان است. بنابراین بسامدهای حرفی، بسامدهای زوجی و غیرهٔ حاصل از دنباله‌های خاص، همچنان‌که طول دنباله‌ها افزایش می‌یابد، به حدودی معین و مستقل از دنبالهٔ خاص نزدیک خواهند شد. درواقع این مطلب برای تمام دنباله‌ها مصداق ندارد اما مجموعه‌ای که این مطلب برای آن نادرست است دارای احتمال صفر می‌باشد. ویژگی ارگودیک تقریباً به معنای همگنی آماری است.



شکل۴ گراف مربوط به منبع مثال پ.


شکل۵ گراف مربوط به منبع مثال ت.

تمام مثالهای زبانهای مصنوعی فوق، ارگودیک هستند. این ویژگی در ارتباط با ساختار گراف مربوطه است. چنانچه این گراف دو ویژگی زیر را دارا باشد [7] در آن صورت فرایند مربوطه ارگودیک می‌شود:

۱. گراف شامل دو قسمت مجزای A و B نباشد به‌طوری که رفتن از رئوس قسمت A به رئوس قسمت B از طریق خطوط گراف در جهت پیکانها، و همچنین رفتن از رئوس قسمت B به‌رئوس قسمت A غیرممکن باشد.

۲. هر زنجیرهٔ بسته از خطوطِ گراف با جهت‌گیری مشابه در پیکانهای خطوط آن، یک «مدار» نامیده می‌شود. «طول» مدار همان تعداد خطوط موجود در آن است. بنابراین در شکل ۵، زنجیرهٔ BEBES یک مدار با طول ۵ است. دومین ویژگی لازم این است که بزرگترین مقسوم علیه مشترک طول مدارات موجود در گراف یک باشد.

اگر شرط اول برقرار باشد اما شرط دوم به خاطر داشتن بزرگترین مقسوم علیه مشترک $d > 1$ نقض گردد، دنباله‌ها دارای نوعی ساختار تناوبی خواهند بود. دنباله‌های متفاوت در d ردهٔ مختلف قرار خواهند گرفت که صرف‌نظر از جابه‌جایی مبدأ (یعنی اینکه کدام حرف رشته، حرف اول نامیده شود) از نظر آماری مشابه‌اند. هر دنباله با جابه‌جا کردن از ۰ تا $d - 1$ می‌تواند از نظر آماری با هر دنبالهٔ دیگری معادل گردد. یک مثال ساده با $d = 2$ بدین قرار است: سه حرف ممکن a و b و c موجودند. بعد از حرف a، حرف b یا c به‌ترتیب با احتمالات ۱/۳ و ۲/۳ می‌آیند. همیشه بعد از هر یک از حروف b یا c حرف a می‌آید. بنابراین یک دنباله نوعی چنین است:

$$a\,b\,a\,c\,a\,c\;a\,c\,a\,b\,a\,c\,a\,b\,a\,b\,a\,c\,a\,c$$

چنین وضعیتی در کار ما اهمیت زیادی ندارد.

اگر شرط اول نقض گردد، گراف مورد نظر را می‌توان به چند زیرگراف که هر یک در شرط اول صدق می‌کنند تجزیه کرد. ما همچنین فرض می‌کنیم که شرط دوم برای هر زیرگراف برقرار باشد. آنچه را در این حالت داریم می‌توان منبع «مختلط» نامید که متشکل از تعدادی مؤلفه‌های خالص است.

مؤلفه‌های متفاوت مربوط به زیرگرافهای متفاوت‌اند. اگر $L_۱$، $L_۲$، $L_۳$، ... منابع مؤلفه‌ای باشند، می‌توانیم بنویسیم:

$$L = p_۱L_۱ + p_۲L_۲ + p_۳L_۳ + \dots$$

که در آن p_i احتمال منبع مؤلفه‌ای L_i است.

از لحاظ فیزیکی چنین وضعیتی به این معنی است که: تعدادی منابع متفاوت $L_۱$، $L_۲$، $L_۳$، ... با ساختار آماری همگن (یعنی ارگودیک) وجود دارند. ما *از پیش* نمی‌دانیم که کدام‌یک به‌کار گرفته خواهند شد، اما همین که دنباله با یک مؤلفهٔ خالص داده‌شدهٔ L_i آغاز شود، مطابق با ساختار آماری آن مؤلفه، به‌طور نامحدود ادامه می‌یابد.

به‌عنوان مثال می‌توان دو فرایند تعریف‌شدهٔ فوق را اختیار کرد و به آنها $p_۱ = ۰٫۲$ و $p_۲ = ۰٫۸$ را نسبت داد. یک دنباله از منبع مختلطِ

$$L = ۰٫۲L_۱ + ۰٫۸L_۲$$

با انتخاب $L_۱$ یا $L_۲$ با احتمالات ۰٫۲ و ۰٫۸ در ابتدا، و پس از آن تولید دنباله از منبع انتخاب‌شده به‌دست می‌آید.

منابع را ارگودیک فرض خواهیم کرد، مگر وقتی‌که خلاف آن ذکر شود. با چنین فرضی می‌توان میانگینها در طول یک دنباله را، با میانگینها روی مجموعهٔ دنباله‌های ممکن، یکی گرفت (احتمال مغایرت صفر است). برای مثال بسامد نسبی حرف A در یک دنبالهٔ نامتناهی مشخص، با احتمالِ یک با بسامدِ نسبیِ آن در مجموع دنباله‌ها برابر خواهد بود.

اگر P_i احتمالِ حالت i و $p_i(j)$ احتمال انتقال به حالت j باشد، برای فرایندی که ساکن و بلاتغییر خواهد شد، کاملاً واضح است که P_i باید در شرایط تعادل صدق کند:

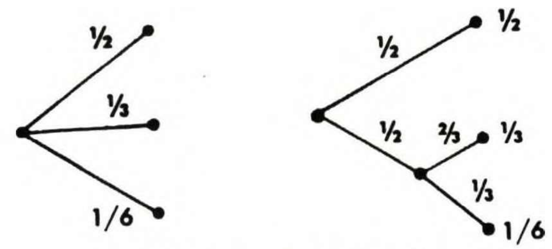
$$P_j = \sum_i P_i p_i(j)$$

در حالتِ ارگودیک می‌توان نشان داد که با هر شرایطِ اولیه‌ای، احتمال بودن در حالتِ j بعد از N نماد، $P_j(N)$، همچنان‌که $N \to \infty$ به مقادیر تعادل نزدیک می‌شود.

۶. انتخاب، عدم حتمیت، آنتروپی

ما منبع اطلاعات گسسته را به‌عنوان فرایند مارکُف بیان کردیم. حال ببینیم که آیا می توانیم کمیتی تعریف کنیم که به طریقی مقدار اطلاعات «تولید شده» توسط چنین فرایندی را اندازه بگیرد یا، بهتر بگوییم، اندازه بگیرد که اطلاعات با چه آهنگی تولید می‌شوند؟

فرض کنید مجموعه‌ای از پیشامدهای ممکن داریم که احتمالات وقوع آنها عبارت‌اند از: $p_۱$، $p_۲$، ...، p_n. البته این احتمالات معلوم‌اند، اما این تمام چیزی است که ما راجع به پیشامدی که قرار است رخ دهد می‌دانیم. آیا ما می‌توانیم شاخصی پیدا کنیم که نشان دهد چه مقدار «انتخاب» در برگزیدن این پیشامد دخیل است و یا اینکه ما راجع به نتیجهٔ آن چقدر تردید داریم؟



شکل ۶ تجزیهٔ یک انتخاب از سه امکان.

چنانچه چنین شاخصی وجود داشته باشد، مثلاً $H(p_۱, p_۲, \dots, p_n)$، کاملاً معقول است که ویژگیهای زیر را از آن بخواهیم:

۱. H باید نسبت به p_i پیوسته باشد.

۲. چنانچه تمام p_iها مساوی باشند، $p_i = ۱/n$، H باید تابعی یکنوای صعودی از n باشد. با چنین پیشامدهای هم‌احتمال، زمانی‌که پیشامدهای ممکنِ بیشتری وجود داشته باشد، انتخاب و عدم حتمیت بیشتری به‌وجود می‌آید.

۳. اگر یک انتخاب را به دو انتخاب متوالی تقسیم کنیم، H اصلی باید همان مجموع وزنیِ مقادیری مجزای H باشد. معنای این مطلب در شکل ۶ توضیح داده شده‌است. در سمت چپ ما سه احتمال $p_۱ = ۱/۲$، $p_۲ = ۱/۳$ و $p_۳ = ۱/۶$ را داریم. در سمت راست ابتدا بین دو حالت ممکن با احتمال ۱/۲ انتخاب می‌کنیم، و اگر دومی رخ داد، بین احتمالات ۲/۳ و ۱/۳ یکی دیگر را برمی‌گزینیم. نتایج نهایی دارای همان احتمالات قبلی‌اند. بنابراین در این موردِ خاص نیازمند آنیم که

$$H\left(\frac{۱}{۲}, \frac{۱}{۳}, \frac{۱}{۶}\right) = H\left(\frac{۱}{۲}, \frac{۱}{۲}\right) + \frac{۱}{۲}H\left(\frac{۲}{۳}, \frac{۱}{۳}\right)$$

ضریب ۱/۲ عامل وزنی مذکور است، زیرا گزینهٔ دومی تنها در نیمی از مواقع رخ می‌دهد.

در پیوست ۲ نتیجهٔ ذیل ثابت شده است:

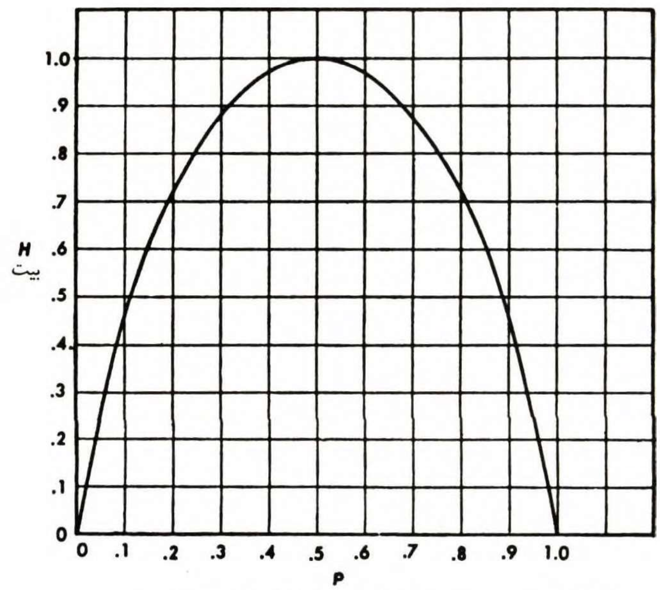
قضیهٔ ۲. *تنها تابع H که در مفروضات سه‌گانهٔ فوق صدق می‌کند به‌صورت زیر است:*

$$H = -K\sum_{i=۱}^{n} p_i \log p_i$$

که در آن K یک عدد ثابت مثبت است.

این قضیه و مفروضات لازم برای اثبات آن به‌هیچ وجه برای ارائهٔ نظریهٔ حاضر لازم نیستند، بلکه عمدتاً برای قابل قبول ساختن بعضی از تعاریف بعدی بیان شده‌اند. به‌هر حال توجیه واقعی این تعاریف در نتایجشان نهفته است.

مقادیری که به‌صورت $H = -\sum p_i \log p_i$ هستند (ثابت K تنها به انتخاب واحد اندازه‌گیری بستگی دارد)، در اندازه‌گیری میزان اطلاعات، انتخاب، و عدم حتمیت نقش اصلی را در نظریهٔ اطلاعات بازی می‌کنند. این صورت H به مثابهٔ فرمول‌بندی مشابهی از آنتروپی در مکانیک آماری



شکل ۷ آنتروپی در مورد دو حالت ممکن با احتمالات p و (p − ۱).

[۸]، جایی‌که p_i احتمال قرار داشتن سیستم در حفرهٔ iام از فضای فاز خود باشد، شناخته می‌شود. در این صورت H، به‌عنوان مثال همان H در قضیهٔ مشهور H بولتسمان است. بنابراین بهتر است ما $H = -\sum p_i \log p_i$ را آنتروپی مجموعهٔ احتمالاتِ $p_1, \ldots, p_n$ بنامیم. اگر x یک متغیر تصادفی باشد، آنتروپیِ آن را به‌صورت $H(x)$ خواهیم نوشت؛ بنابراین x شناسهٔ تابع نیست بلکه برچسبی برای یک عدد است که آن را از $H(y)$ یا آنتروپی متغیر تصادفی y متمایز می‌کند.

آنتروپی در مورد دو حالت ممکن با احتمالاتِ p و $q = 1 - p$، یعنی

$$H = -(p \log p + q \log q)$$

در شکل ۷ به‌عنوان تابعی از p ترسیم شده است.

کمیّت H خواص جالبی دارد که باعث می‌شود شاخص معقولی برای اندازه‌گیری انتخاب یا اطلاعات باشد.

۱. $H = 0$ اگر و فقط اگر تمام p_i ها صفر باشند بجز یکی که مقدارش برابر واحد است. بنابراین H تنها زمانی که ما کاملاً به‌نتیجه یقین داریم صفر می‌شود و در غیراین صورت‌مثبت است.

۲. به ازای nای داده شده، H هنگامی که همهٔ p_i ها باهم برابر، یعنی مساوی $1/n$ باشند، بیشینه و برابر با $\log n$ خواهد بود. این وضعیت به‌طور شهودی نیز نامشخص‌ترین وضعیت است.

۳. فرض کنید در مسألهای دو پیشامد x و y با m حالت ممکن برای اولی و n حالت برای دومی موجودند، و $p(i,j)$ احتمال وقوع توأم i برای اولی و j برای دومی باشد. آنتروپیِ پیشامد توأم برابر

$$H(x,y) = \sum_{i,j} p(i,j) \log p(i,j)$$

می‌باشد، و

$$H(x) = -\sum_{i,j} p(i,j) \log \sum_{j} p(i,j)$$

$$H(y) = -\sum_{i,j} p(i,j) \log \sum_{i} p(i,j)$$

به‌سادگی می‌توان نشان داد که

$$H(x,y) \le H(x) + H(y)$$

که تساوی تنها در صورت مستقل بودن پیشامدهاست (یعنی $p(i,j) = p(i)p(j)$). عدم حتمیت یک پیشامدِ توأم، کوچکتر یا مساوی مجموع عدم حتمیتهای مجزاست.

۴. هر تغییری در جهت متعادل ساختن احتمالات $p_1, p_2, \ldots, p_n$ مقدار H را افزایش می‌دهد. بنابراین اگر $p_1 < p_2$ و ما p_1 را افزایش دهیم، و p_2 را به همان میزان کاهش دهیم تا p_1 و p_2 تقریباً مساوی شوند، H افزایش خواهد یافت. به‌عبارت کلیتر، اگر هر نوع عمل «میانگین‌گیری» به‌صورت زیر روی p_iها اعمال کنیم

$$p'_i = \sum_j a_{ij} p_j$$

به‌طوری که

$$\sum_i a_{ij} = \sum_j a_{ij} = 1$$

و تمام a_{ij}ها نامنفی باشند، در آن صورت H افزایش می‌یابد (به استثنای مورد خاصی که این تبدیل چیزی بیش از یک جایگشتِ p_jها نباشد، که البته در آن صورت H تغییری نمی‌کند).

۵. مانند مورد ۳ فرض کنید دو پیشامد تصادفی x و y موجودند که لزوماً مستقل نیستند. برای هر مقدار مشخص i که x بتواند اختیار کند یک احتمال شرطی $p_i(j)$، یعنی احتمالِ اینکه y مقدار j را بگیرد، وجود دارد و به‌صورت زیر به‌دست می‌آید

$$p_i(j) = \frac{p(i,j)}{\sum_j p(i,j)}$$

آنتروپی شرطی y، $H_x(y)$، را به‌صورت میانگین آنتروپی y برای هر مقدار x تعریف می‌کنیم که متناظر با احتمال آن مقدار خاص x، وزنی شده است. یعنی

$$H_x(y) = -\sum_{i,j} p(i,j) \log p_i(j)$$

این کمیت، میانگین میزان تردید ما را در مورد y هنگامی که x را می‌دانیم اندازه‌گیری می‌کند. با جایگزینی مقدار $p_i(j)$ داریم:

$$\begin{aligned} H_x(y) &= -\sum_{i,j} p(i,j) \log p(i,j) + \sum_{i,j} p(i,j) \log \sum_j p(i,j) \\ &= H(x,y) - H(x) \end{aligned}$$

یا

$$H(x,y) = H(x) + H_x(y)$$

عدم حتمیت (یا آنتروپی) پیشامدِ توأم x و y برابر است با عدم حتمیت x به‌علاوهٔ عدم حتمیت y، زمانی که x معلوم است.

۶. از ۳ و ۵ داریم

$$H(x) + H(y) \geq H(x,y) = H(x) + H_x(y)$$

بنابراین

$$H(y) \geq H_x(y)$$

عدم حتمیت y با دانستن x هرگز افزایش نمی‌یابد، و بجز وقتی که x و y پیشامدهای مستقلی باشند، که در چنان وضعی این عدم حتمیت تغییری نمی‌کند، کاهش خواهد یافت.

۷. آنتروپی یک منبع اطلاعات

منبعی گسسته از نوع حالات متناهی را که در بالا شرح داده شد در نظر بگیرید. به ازای هر حالت ممکن i مجموعه‌ای از احتمالات $p_i(j)$ برای تولید نمادهای مختلف ممکن j وجود خواهد داشت. بنابراین برای هر حالت i یک آنتروپیِ H_i موجود است. آنتروپی منبع را می‌توان به‌صورت میانگین H_iها که برحسب احتمال وقوع حالتهای موردنظر وزندار شده‌اند، تعریف کرد.

$$\begin{aligned} H &= \sum_j P_i H_j \\ &= -\sum_{ij} P_i p_i(j) \log p_i(j) \end{aligned}$$

این عبارت، آنتروپی منبع بر نمادِ متن است. اگر این فرایندِ مارکُفْ نمادها را با آهنگ زمانیِ مشخصی تولید کند، یک آنتروپی بر ثانیه هم موجود است:

$$H' = \sum_i f_i H_i$$

که در آن f_i میانگین بسامد (رخدادها در ثانیه) در حالت i است. به‌روشنی داریم

$$H' = mH$$

که m میانگین تعداد نمادهای تولید شده در ثانیه است. H یا H' مقدار اطلاعات تولیدی منبع بر نماد یا بر ثانیه را اندازه‌گیری می‌کنند. اگر پایهٔ لگاریتم ۲ باشد، آنها تعداد بیتها بر نماد یا بر ثانیه را بیان خواهند نمود.

چنانچه نمادهای متوالی مستقل باشند، H به‌سادگی برابر $-\sum p_i \log p_i$ است که در آن p_i احتمال نماد i می‌باشد. در این مورد فرض کنید یک پیام بلندِ N نمادی را در نظر گرفته‌ایم. این پیام به احتمال زیاد شامل حدوداً $p_1 N$ مورد از نماد اول، $p_2 N$ مورد از نماد دوم و غیره خواهد بود. بنابراین احتمالِ این پیامِ مشخص تقریبا برابر خواهد بود با

$$p = p_1^{p_1 N}\, p_2^{p_2 N} \dots p_n^{p_n N}$$

یا

$$\log p \approx N \sum_i p_i \log p_i$$

$$\log p \approx -NH$$

$$H \approx \frac{\log \frac{1}{p}}{N}$$

بنابراین H تقریباً برابر لگاریتم احتمال معکوس یک دنبالهٔ بلند نوعی، تقسیم بر تعداد نمادهای موجود در آن دنباله می‌باشد. نتیجهٔ مشابهی برای هر نوع منبعی برقرار است. اگر بخواهیم دقیقتر بگوییم، داریم (نگاه کنید به پیوست ۳):

قضیهٔ ۳. *برای هر $\epsilon > 0$ و $\delta > 0$ داده شده، می‌توانیم N_0ای بیابیم که دنباله‌های به هر طولِ $N \geq N_0$ در یکی از دو دستهٔ زیر قرار بگیرند:*

۱. مجموعه‌ای که احتمال کلی آن کوچکتر از ϵ است.

۲. بقیه، یعنی تمام آنهایی که اعضایشان احتمالاتی دارند که در نابرابری زیر صدق می‌کنند:

$$\left| \frac{\log p^{-1}}{N} - H \right| < \delta$$

به‌عبارت دیگر ما تقریباً یقین داریم که زمانی که N بزرگ باشد، $(\log p^{-1})/N$ خیلی به H نزدیک است.

قضیه‌ای بسیار مرتبط با این قضیه در مورد دنباله‌هایِ با احتمالات گوناگون وجود دارد. یک بار دیگر دنباله‌هایی با طول N در نظر بگیرید که با ترتیبِ نزولیِ احتمال قرار گرفته‌اند. $n(q)$ را به‌منزلهٔ تعدادی تعریف می‌کنیم که باید از این مجموعه برداریم تا با شروع از محتملترین عضو و انباشتن احتمال آنها، به احتمال مجموع q برسیم.

قضیهٔ ۴

$$\lim_{N \to \infty} \frac{\log n(q)}{N} = H$$

زمانی که q برابر با صفر یا یک نباشد.

می‌توانیم $\log n(q)$ را به‌عنوان تعداد بیتهای لازم برای مشخص کردن یک دنباله، زمانی که فقط محتملترین دنباله‌ها را با مجموع احتمال q در نظر گرفته‌ایم، تعبیر کنیم. در این صورت $[\log n(q)]/N$ تعداد بیت بر نماد برای این مشخص‌سازی است. قضیهٔ مذکور به ما می‌گوید که برای Nهای بزرگ، این عبارت، مستقل از q، برابر H خواهد بود. آهنگ رشدِ لگاریتمِ تعدادِ دنباله‌های بسیار محتمل به‌وسیلهٔ H داده می‌شود، صرف‌نظر از اینکه تعبیر ما از «بسیار محتمل» چه باشد. بنا بر این نتایج که در پیوست ۳ اثبات

شده‌اند، برای بیشتر مقاصد این امکان وجود دارد که با دنباله‌های بلند طوری رفتار کنیم که گویی تعدادشان تنها 2^{HN} و احتمال هر یک 2^{-HN} است.

دو قضیهٔ بعدی نشان می‌دهند که H و H' با انجام اعمال حدگیری مستقیماً از آمارِ دنباله‌های پیام و بدون رجوع به حالات و احتمالاتِ انتقالی بین حالات مشخص می‌گردند.

قضیهٔ ۵. *$p(B_i)$ را احتمال یک دنبالهٔ B_i از نمادهایی از منبع در نظر بگیرید. همچنین*

$$G_N = -\frac{1}{N}\sum_i p(B_i)\log p(B_i)$$

که مجموع، روی کلیهٔ دنباله‌های B_i که شامل N نماد هستند می‌باشد. در این صورت G_N یک تابع یکنوای نزولی از N است و داریم:

$$\lim_{N\to\infty} G_N = H$$

قضیهٔ ۶. *$p(B_i, S_j)$ را احتمال دنبالهٔ B_i که نماد S_j را به دنبال دارد، و $p_{B_i}(S_j) = p(B_i, S_j)/p(B_i)$ را احتمال شرطیِ ظهور S_j بعد از B_i در نظر بگیرید. فرض کنید*

$$F_N = -\sum_{i,j} p(B_i, S_j)\log p_{B_i}(S_j)$$

که مجموع، روی کلیهٔ دسته‌های $N-1$ نمادیِ B_i و کلیهٔ نمادهای S_j است. در این صورت F_N یک تابع یکنوای نزولی از N است و

$$F_N = NG_N - (N-1)G_{N-1}$$

$$G_N = \frac{1}{N}\sum_{1}^{N} F_N$$

$$F_N \le G_N$$

و

$$\lim_{N\to\infty} F_N = H$$

این نتایج در پیوست ۳ استنتاج شده‌اند، و نشان می‌دهند که تقریبهایی از H را می‌توان تنها با در نظر گرفتن ساختار آماری دنباله‌هایی که روی ۱، ۲، ...، N نماد بسط داده شده‌اند به‌دست آورد. F_N تقریب بهتری است و در واقع F_N آنتروپی تقریبیِ مرتبهٔ Nام منبعی از نوع بالاست. اگر تأثیر آماری در بسط بیش از N نماد وجود نداشته باشد ــ که این خود یعنی اینکه احتمال شرطیِ نماد بعدی در حالی که $(N-1)$ نماد قبلی را می‌دانیم، با دانستن هر تعدادی قبل از آن هیچ تغییری نکند ــ آنگاه $F_N = H$. البته F_N آنتروپی شرطیِ نماد بعدی به شرط دانستن $(N-1)$ تای قبلی است در حالی‌که G_N آنتروپی بر نماد دسته‌های N نمادی است.

نسبت آنتروپی منبع به حداکثر مقداری که با محدود ماندن به همان نمادها می‌توانیم داشته باشیم *آنتروپی نسبی* نامیده می‌شود. این نسبت، همان‌گونه که بعداً خواهیم دید، وقتی که از همان الفبا برای کدگذاری استفاده شود بالاترین حد ممکن فشردگی است. یک منهای آنتروپی نسبی را حشو[1] می‌نامند.

1. redundancy

حشوِ زبان انگلیسیِ معمولی، قطع نظر از ساختار آماری روی فواصلی با بیش از هشت حرف، حدوداً ۵۰ درصد است. این بدان معناست که وقتی ما انگلیسی می‌نویسیم، نیمی از آنچه می‌نویسیم توسط ساختار زبان تعیین و نیم دیگر آن آزادانه انتخاب می‌شود. عدد ۵۰ درصد به‌وسیلهٔ روشهای متعدد مستقلی به‌دست آمده است که همهٔ آنها نتایجی در همین حدود را به‌دست می‌دهند. یکی از این روشها از طریق محاسبهٔ آنتروپیِ تقریبهای زبان انگلیسی است. دومین روش این است که کسر خاصی از حروفِ یک نمونهٔ متن انگلیسی را حذف کنند و سپس شخصی سعی کند دوباره آنها را برگرداند. چنانچه آن شخص بتواند در حالی‌که ۵۰ درصد حروف حذف شده‌اند آنها را باز گرداند، حشو باید بیش از ۵۰ درصد باشد. سومین روش در ارتباط با یک نتیجهٔ شناخته شده در رمزنگاری است.

دو مُنتهای حشو در نثر انگلیسی، در کتابِ *Basic English* و کتاب *Finnegans Wake* اثر جیمز جویس ارائه شده‌اند. تعداد کلمات کتابِ *Basic English* به ۸۵۰ لغت محدود می‌شود و حشو آن خیلی زیاد است. این مطلب در طول و تفصیلی که در هنگام برگردان قطعه متنی به نثر *Basic English* رخ می‌دهد، آشکار می‌شود. در سوی دیگر، جویس مجموعهٔ لغات را گسترش داده و گفته می‌شود که به یک فشردگی محتوای معنایی دست یافته است.

حشو یک زبان با وجود جداول کلمات متقاطع مرتبط است. چنانچه حشو صفر باشد هر دنباله‌ای از حروف، یک متن معقول در زبان محسوب می‌شود و هر آرایهٔ دوبعدی از حروف یک جدول کلمات متقاطع را تشکیل می‌دهد. اگر حشو خیلی زیاد باشد، زبان محدودیتهای بسیاری بر روی جداول کلمات متقاطع ممکن اعمال می‌کند. یک تحلیل مفصلتر نشان می‌دهد که چنانچه فرض کنیم محدودیتهای اعمال شده توسط زبان نسبتاً آشفته و تصادفی باشند، آن وقت جداول کلمات متقاطع بزرگ تنها زمانی ممکن خواهند بود که حشو ۵۰ درصد باشد. اگر حشو ۳۳ درصد باشد جداول کلمات متقاطع سه‌بعدی امکان پذیر خواهند شد و غیره.

۸. معرفی اعمال کدگذاری و کُدگشایی[1]

هنوز اعمالی را که توسط فرستنده و گیرنده در کدگذاری و کُدگشایی اطلاعات صورت می‌گیرد، به‌صورت ریاضی بیان نکرده‌ایم. هر دوی اینها تراگردان[2] گسسته نامیده می‌شوند. ورودیِ یک تراگردان دنباله‌ای از نمادهای ورودی و خروجی آن دنباله‌ای از نمادهای خروجی است. تراگردان ممکن است حافظهٔ داخلی هم داشته باشد که در این صورت خروجی آن نه تنها به نماد ورودی بلکه به سابقهٔ گذشتهٔ آن نیز بستگی دارد. ما حافظهٔ داخلی را متناهی فرض می‌کنیم، یعنی تعداد متناهیِ m عدد از حالتهای ممکن برای تراگردان موجود است و خروجی، تابعی از حالت فعلی و نماد ورودی فعلی می‌باشد. حالت بعدیْ تابع دیگری از همین دو کمیت خواهد بود. بنابراین یک تراگردان را می‌توان با دو تابع توصیف کرد:

$$y_n = f(x_n, \alpha_n)$$

$$\alpha_{n+1} = g(x_n, \alpha_n)$$

که در آن x_n عبارت است از nامین نماد ورودی، α_n حالت تراگردان هنگام

1. decoding 2. transducer

دیدن nامین نماد ورودی، y_n نماد خروجی (یا دنبالهٔ نمادهای خروجی) تولید شده است وقتی که حالت تراگردان α_n باشد و x_n دیده شود.

اگر نمادهای خروجی تراگردانی را بتوان به‌عنوان نمادهای ورودیِ تراگردان دیگری شناخت، آنها را می‌توان پشت سر هم قرار داد و حاصل کماکان یک تراگردان خواهد بود. چنانچه تراگردان دومی موجود باشد که بر روی خروجیِ اولی عمل کرده و ورودی اولیه را برگرداند، تولید کنندهٔ اول را ناتکین و دومی را وارون آن خواهیم نامید.

قضیهٔ ۷. *خروجیِ یک تراگردان با حالات متناهی که از یک منبعِ آماری متناهی تغذیه می‌شود، خود یک منبع آماری متناهی است و آنتروپی آن (در هر لحظه) کوچکتر یا مساوی آنتروپی ورودی است، که اگر تراگردان ناتکین باشد این دو مساوی خواهند بود.*

فرض کنید α بیانگر حالت منبعی است که دنبالهٔ نمادهای x_i را تولید کرده و β حالت تراگردانی است که در خروجیش دسته‌های نمادهای y_j را تولید نموده است. سیستم ترکیبی می‌تواند به وسیلهٔ «فضای حالتِ حاصلضربی» از زوجهای (α, β) ارائه شود. دو نقطه در فضا، (α_1, β_1) و (α_2, β_2)، با یک خط به یکدیگر متصل‌اند به شرط آنکه α_1 بتواند xای تولید کند که β_1 را به β_2 تغییر دهد و در این صورت، به این خط احتمال آن x داده می‌شود. این خط با دسته نمادهای y_1 که به‌وسیلهٔ تراگردان تولید می‌شود، علامت‌گذاری می‌گردد. آنتروپی خروجی را می‌توان به‌عنوان مجموع وزنی روی حالات محاسبه کرد. اگر مجموع را ابتدا روی β بگیریم هر عبارتِ حاصل کوچکتر یا مساوی عبارت مربوط برای α است، بنابراین آنتروپی افزایش نمی‌یابد. اگر تراگردان ناتکین باشد فرض می‌کنیم خروجیش به تراگردانِ وارون متصل می‌گردد. حال اگر H'_1، H'_2 و H'_3 به‌ترتیب آنتروپیِ منبع، تراگردانِ اولی و دومی باشند، آنگاه $H'_1 \geq H'_2 \geq H'_3 = H'_1$ و بنابراین $H'_1 = H'_2$.

فرض کنید سیستمی از قیود بر روی دنباله‌های ممکن داریم که می‌توان آن را به‌وسیلهٔ گرافِ خطیِ شکل ۲ نشان داد. اگر احتمالات $p_{ij}^{(s)}$ را به خطوط مختلفی نسبت دهیم که حالت i را به حالت j متصل می کنند، به یک منبع می‌رسیم. تخصیص مشخصی وجود دارد که آنتروپی حاصل را بیشینه می‌سازد (پیوست ۴ [این پیوست در اینجا نیامده است]).

قضیهٔ ۸. *فرض کنید سیستمی از قیود را به‌عنوان یک کانال با ظرفیت $C = \log W$ در نظر بگیریم. چنانچه قرار دهیم*

$$p_{ij}^{(s)} = \frac{B_j}{B_i} W^{-l_{ij}^{(s)}}$$

که در آن $l_{ij}^{(s)}$ طول مدت sامین نمادی است که حالت i را به حالت j هدایت می‌کند و B_i در تساوی زیر صدق می‌کند

$$B_i = \sum_{s,j} B_j W^{-l_{ij}^{(s)}}$$

در آن صورت H بیشینه و برابر C خواهد شد.

با تخصیص صحیح احتمالاتِ انتقالْ آنتروپی نمادها بر روی کانال می‌تواند تا حد ظرفیت کانال بیشینه شود.

۹. قضیهٔ اساسی کانال بدون نویز

اکنون می‌خواهیم تعبیر خودمان از H را به‌عنوان آهنگ تولید اطلاعات، با اثبات این مطلب که H ظرفیت لازم برای کانال را در مؤثرترین حالت کُدگذاری مشخص می‌کند، توجیه نماییم.

قضیهٔ ۹. *فرض کنید منبعی دارای آنتروپی H (بیت بر نماد) و کانالی دارای ظرفیت C (بیت بر ثانیه) است. این امکان وجود دارد که خروجی منبع را به روشی کدگذاری کنیم که بتواند با میانگین آهنگِ $C/H - \epsilon$ نماد بر ثانیه روی کانال فرستاده شود، که ϵ به دلخواه کوچک است، و غیرممکن است که انتقال با میانگین آهنگی بیش از C/H انجام پذیرد.*

عکس این قضیه، یعنی اینکه نمی‌توان از C/H فراتر رفت، می‌تواند با توجه به این اثبات شود که آنتروپی وارده به کانال در ثانیه با همان مقدار منبع برابر است، زیرا کانال فرستنده باید ناتکین باشد، و نیز اینکه این آنتروپی نمی‌تواند بیشتر از ظرفیت باشد. بنابراین $H' \leq C$ و تعداد نمادها در ثانیه برابر $H'/H \leq C/H$ است.

بخش اول قضیه با دو روش مختلف اثبات خواهد شد. روش اول در نظر گرفتن مجموعهٔ تمام دنباله‌های N نمادی است که توسط منبع تولید می‌شوند. برای N بزرگ می‌توانیم آنها را به دوگروه تقسیم کنیم که اولی شامل کمتر از $2^{(H+\eta)N}$، و دومی شامل کمتر از 2^{RN} عضو با احتمال مجموع کمتر از μ است (که R لگاریتم تعداد نمادهای متفاوت است). همچنان که N افزایش می‌یابد μ و η به صفر میل می‌کنند. تعداد سیگنالهای با طول مدت T در کانال بزرگتر است از $2^{(C-\theta)T}$ که وقتی T بزرگ باشد، θ کوچک خواهد بود. با انتخابِ

$$T = \left(\frac{H}{C} + \lambda\right) N$$

وقتی که N و T به حد کافی بزرگ باشند (λ هر چقدر هم کوچک باشد) تعداد کافی از دنباله‌های نمادهای کانال برای گروه محتملتر به همراه تعدادی علاوه بر آن وجود خواهند داشت. گروه محتملتر با روش یک به یکِ دلخواهی با اعضای این مجموعه کدگذاری می‌شوند. دنباله‌های باقیمانده [منبع] به‌وسیلهٔ دنباله‌های بلندتر [کانال]، که با یکی از دنباله‌های استفاده نشده [کانال] برای گروه محتملتر شروع و ختم می‌شوند، کدگذاری می‌گردند. این دنباله‌های خاص [کانال] به‌عنوان سیگنال شروع و خاتمهٔ کُدهای دیگر عمل می‌کنند. در ضمن برای دیگر دنباله‌های لازم برای کلیهٔ پیامهای کم‌احتمال نیز، وقت کافی باید در نظر گرفته شود. این مستلزم

$$T_1 = \left(\frac{R}{C} + \varphi\right) N$$

است، که در آن φ کوچک است. در این صورت میانگینِ آهنگ انتقال برحسب نماد بر ثانیه بزرگتر خواهد بود از

$$\left[(1-\delta)\frac{T}{N} + \delta\frac{T_1}{N}\right]^{-1} = \left[(1-\delta)\left(\frac{H}{C} + \lambda\right) + \delta\left(\frac{R}{C} + \varphi\right)\right]^{-1}$$

همچنان که N افزایش می‌یابد، δ، λ و φ به سمت صفر، و آهنگ مذکور به C/H میل می‌کنند.

روش دیگر انجام این کدگذاری و در نتیجه اثبات قضیه را می‌توان به‌صورت ذیل شرح داد: پیامهای با طول N را به‌ترتیب نزولی احتمال مرتب کرده، فرض می‌کنیم که احتمالاتشان $p_1 \geq p_2 \geq p_3 \ldots \geq p_n$ باشد. همچنین $P_s = \sum_1^{s-1} p_i$ که در آن P_s احتمال تجمعی تا (و نه شاملِ) p_s است. ما ابتدا کدگذاری را در یک سیستم دودویی انجام می‌دهیم. کُدِ دودویی پیام از بسط P_s به‌عنوان یک عدد مبنای دو به‌دست می‌آید. این بسط در m_s رقم انجام می‌شود، که m_s عددی است صحیح که در نامساوی زیر صدق می‌کند

$$\log_2 \frac{1}{p_s} \leq m_s < 1 + \log_2 \frac{1}{p_s}$$

بنابراین پیامهای با احتمالِ زیاد به وسیلهٔ کُدهای کوتاهتر و آنهایی که احتمال کمتری دارند با کُدهای بلندتر بیان می‌شوند. از این نابرابریها داریم

$$\frac{1}{2^{m_s}} \leq p_s < \frac{1}{2^{m-1}}$$

کُد P_s با کلیهٔ آنها که به دنبالش قرار گرفته‌اند در یک یا چند رقم، از m_s رقم آن متفاوت است، چرا که کلیهٔ P_iهای باقیمانده دست‌کم از $1/2^{m_s}$ بزرگترند و از این رو بسط دودویی آنها در m_s رقم اول متفاوت است. در نتیجه کلیهٔ کُدها متفاوت‌اند و بازیافت پیام از روی کُد آن امکان‌پذیر است. اگر دنباله‌های کانال از ابتدا دنباله‌های ارقام دودویی نباشند، می‌توان آنها را به‌طور قراردادی به اعداد دودویی منتسب کرد و سپس کُدهای دودویی را به سیگنالهای مناسب برای کانال ترجمه نمود. میانگین تعداد رقمهای دودویی استفاده شده در هر نماد پیام اولیه، H_1، به سادگی برآورد می‌گردد. داریم

$$H_1 = \frac{1}{N} \sum m_s p_s$$

اما

$$\frac{1}{N} \sum \left(\log_2 \frac{1}{p_s}\right) p_s \leq \frac{1}{N} \sum m_s p_s < \frac{1}{N} \sum \left(1 + \log_2 \frac{1}{p_s}\right) p_s$$

و در نتیجه

$$G_N \leq H_1 < G_N + \frac{1}{N}$$

همچنان که N افزایش می‌یابد G_N به H میل می‌کند، آنتروپی منبع و H_1 نیز به H میل می‌کنند.

از اینجا می‌بینیم که غیرمؤثر بودن عمل کُدگذاری هنگامی که تنها از تأخیر متناهی N نمادی استفاده می‌شود، لزوماً بیش از $1/N$ به‌علاوهٔ مابه‌التفاوت آنتروپی واقعیِ H و مقدار محاسبه شدهٔ G_N برای دنباله‌های N نمادی نخواهد بود. و بنابراین درصد زمان لازم مازاد بر ایده‌آل کمتر است از

$$\frac{G_N}{H} + \frac{1}{HN} - 1$$

این روشِ کدگذاری اساساً مشابه روشی است که مستقلاً به وسیلهٔ فانو [۹] پیدا شده است. روش وی قراردادن پیامهای با طول N به‌ترتیب نزولیِ احتمال و تقسیم این زنجیره به دو گروه با احتمالهایِ تا حد ممکنْ مساوی است. اگر پیام در گروه اول باشد اولین رقم دودویی آن صفر و در غیر این صورت یک خواهد بود. این گروهها به‌طور مشابه به زیرمجموعه‌های تقریباً هم‌احتمال تقسیم می‌شوند و زیرمجموعهٔ مربوط، دومین رقم دودویی را مشخص می‌سازد. این فرایند تا زمانی که هر زیرمجموعه تنها شامل یک پیام باشد، ادامه می‌یابد. به‌سادگی دیده می‌شود که به استثنای تفاوتی جزئی (معمولاً در رقم آخر)، به همان چیزی که در فرایند حسابی فوق تشریح شد، خواهیم رسید.

۱۰. بحث و مثال

برای اینکه بتوانیم بیشترین انتقال قدرت از یک ژنراتور به بار را به‌دست آوریم، معمولاً باید از یک مبدل[1] استفاده کنیم تا ژنراتور از دید بار، مقاومت بار را داشته باشد. موقعیتْ در بحثِ ما تقریباً مشابهِ این است. تراگردانی که عمل کُدگذاری را انجام می‌دهد باید از لحاظ آماری منبع را با کانال تطبیق دهد. این منبع از دید تراگردانی که از طریق کانال به آن می‌نگرد، باید دارای ساختار آماری مشابهی همچون منبعی باشد که آنتروپی کانال را بیشینه می‌سازد. مضمون قضیهٔ ۹ این است که اگرچه به‌طور کلی مطابقتِ محض ممکن نیست ولی ما می‌توانیم با بیشترین دقتِ مطلوب آن را تقریب بزنیم. نسبت میزان واقعیِ انتقال به ظرفیت C را می‌توان کارایی سیستم کُدگذاری نامید. که البته این برابر با نسبت آنتروپیِ واقعیِ نمادهای کانال به حداکثر آنتروپی ممکن است.

به‌طور کلی، کدگذاریِ مطلوب یا نسبتاً مطلوب مستلزم تأخیری طولانی در فرستنده و گیرنده است. در وضعیتِ بدون نویز که پیش از این بررسی کردیم، نقش اصلیِ این تأخیر به‌وجود آوردن هماهنگی مناسبی بین احتمالات و طول دنباله‌های مربوطه است. با یک کُد خوب، لگاریتمِ معکوس احتمالِ یک پیام طولانی باید با مدت سیگنال مربوطه متناسب باشد، در حقیقت

$$\left| \frac{\log p^{-1}}{T} - C \right|$$

باید برای کلیهٔ پیامهای طولانی، به استثنای کسر کوچکی از آنها، کوچک باشد.

اگر یک منبع، تنها بتواند یک پیام خاص را تولید کند، آنتروپی آن صفر است و کانالی لازم نیست. به‌عنوان مثال، ماشین محاسبی که برای محاسبهٔ ارقام متوالیِ π ساخته شده است دنبالهٔ مشخصی را بدون هیچ عنصر تصادفی تولید می‌کند. برای «انتقال» آن به نقطه‌ای دیگر به کانالی احتیاج نداریم. شخصی دیگر می‌تواند ماشین دومی را در همان نقطه برای محاسبهٔ همان دنباله مهیا سازد. با این حال، ممکن است این کار غیرعملی به نظر برسد. در چنین موردی می‌توانیم به اختیار، برخی یا کلِ دانش آماریی را که از منبع داریم نادیده بگیریم. می‌توانیم ارقام π را به‌صورت یک دنبالهٔ تصادفی در نظر بگیریم و بر این اساس سیستمی که قادر به ارسال همهٔ دنباله‌های ارقام باشد ایجاد کنیم. به روشی مشابه هم می‌توانیم با استفاده از بخشی از

1. transformer

دانش آماری خود در زبان انگلیسی، و نه همهٔ آن، کُدی بسازیم. در چنین حالتی، منبع را با حداکثر آنتروپی در نظر می‌گیریم مشروط بر اینکه شرایط آماری مورد نظر ما حفظ شود. آنتروپی این منبع، ظرفیت لازم و کافیِ کانال را تعیین می‌نماید. در مثالِ π تنها اطلاعِ حفظ شده این است که تمام ارقام از مجموعهٔ ۰، ۱، ...، ۹ انتخاب شده‌اند. در مورد زبان انگلیسی ممکن است شخصی بخواهد از امکان صرفه‌جوییِ آماری بنابر بسامدهای حرفی استفاده کند، و لاغیر. در این صورت آنتروپیِ بیشینه، اولین تقریب انگلیسی است و آنتروپی این تقریب ظرفیت کانال مورد لزوم را تعیین می‌کند.

به‌عنوان مثالی ساده از برخی از این نتایج، منبعی را در نظر بگیرید که دنباله‌ای از حروف منتخب از میان A، B، C، D با احتمالات ۱/۲، ۱/۴، ۱/۸، ۱/۸ را تولید می‌کند که در آن نمادهای متوالی به‌صورت مستقل انتخاب شده‌اند. داریم

$$H = -(\frac{۱}{۲}\log\frac{۱}{۲} + \frac{۱}{۴}\log\frac{۱}{۴} + \frac{۲}{۸}\log\frac{۱}{۸})$$
$$= \frac{۷}{۴} \text{ بیت بر نماد}$$

بر این اساس می‌توانیم یک سیستم کُدگذاری را تقریب بزنیم و پیامها را از این منبع به ارقام دودویی با میانگین ۷/۴ رقم دودویی برای هر نماد، کدگذاری کنیم. در این مورد واقعاً می‌توانیم با کُدهای ذیل (برگرفته از دومین اثبات قضیهٔ ۹) به این مقدار حدی برسیم:

A	۰
B	۱۰
C	۱۱۰
D	۱۱۱

شمار متوسط اعداد دودویی استفاده شده در کُدگذاریِ یک دنبالهٔ N نمادی برابر خواهد بود با

$$N(\frac{۱}{۲}\times ۱ + \frac{۱}{۴}\times ۲ + \frac{۲}{۸}\times ۳) = \frac{۷}{۴}N$$

به‌سادگی می‌توان مشاهده کرد که ارقام دودوییِ صفر و یک دارای احتمالاتِ ۱/۲، ۱/۲ هستند، بنابراین H برای دنباله‌های کُدگذاری شده یک بیت در هر نماد است. از آنجایی که ما به‌طور متوسط برای هر حرف اصلی، ۷/۴ نماد دودویی داریم، آنتروپی‌های زمانی آنها یکسان هستند. حداکثر آنتروپی ممکن برای مجموعهٔ اصلی عبارت است از $\log ۴ = ۲$، و آن زمانی رخ می‌دهد که A، B، C، D دارای احتمالاتِ ۱/۴، ۱/۴، ۱/۴، ۱/۴، باشند. بنابراین آنتروپی نسبی ۷/۸ است. می‌توانیم از طریق جدول ذیلْ دنباله‌های دودویی را به‌صورت دو به یک به مجموعهٔ اصلی نمادها برگردانیم

۰۰	A'
۰۱	B'
۱۰	C'
11	D'

این فرایند مضاعف، سپس پیام اصلی را به نمادهای مشابه کدگذاری می‌کند، اما نسبت فشردگی متوسط آن ۷/۸ است.

به‌عنوان مثال دوم منبعی را در نظر بگیرید که دنباله‌ای از Aها و Bها را با احتمالاتِ p برای A و q برای B تولید می‌کند. حال اگر $p \le q$ داریم:

$$H = -\log p^p(۱-p)^{۱-p}$$
$$= -p\log p(۱-p)^{(۱-p)/p}$$
$$= p\log\frac{e}{p}$$

در چنین حالتی فرد می‌تواند در یک کانال صفر و یک، با ارسال یک دنبالهٔ خاص، پیام را به‌صورت نسبتاً خوبی کدگذاری کند، مثلاً ۰۰۰۰ برای نماد نامتداول A و در پی آن دنباله‌ای نشانگر *تعداد* Bهایی که بعد از آن می آیند. این را می‌توان از طریق بیان دودوییِ کلیهٔ اعداد نشان داد که در آنها دنبالهٔ خاصی حذف شده است. تمام اعداد تا ۱۶ مثل معمول ارائه می‌گردند؛ خود عدد ۱۶ از طریق عدد دودوییِ بعدیِ ۱۶ که شامل ۰۰۰۰ نمی‌باشد، یعنی ۱۷ = ۱۰۰۰۱، بیان می‌شود و غیره.

در واقع می‌توان نشان داد که وقتی $p \to ۰$، کُدگذاری نیز به سمت ایده‌آل نزدیک می‌شود، به شرط آنکه طول دنبالهٔ خاص به درستی تنظیم شود.

مراجع

1. Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.

2. Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

۳. برای نمونه، رک.

S. Chandrasekhar, "Stochastic Problems in Physics and Astronomy," *Reviews of Modern Physics*, v. 15, No. 1. January 1943, p. 1.

4. Kendall and Smith, *Tables of Random Sampling Numbers*, Cambridg, 1939.

۵. جداول بسامدهای حروف، زوجها، و سه‌تاییها در

Secret and Urgent by Fletcher Pratt, Blue Ribbon Books, 1939

و جدولهای بسامد کلمات در

Relative Frequency of English Speech Sounds, G. Dewey, Harvard University Press, 1923.

آمده است.

۶. برای ملاحظهٔ بحث مبسوطی در این زمینه، رک.

M. Frechet, *Methods des fonctions arbitraires. Theorie des énénements en chaine dans le cas d'un nombre fini d'états possibles.* Paris, Gauthier Villars, 1938.

۷. این گزاره‌ها بیان شرایط ارائه‌شده در مرجع قبل برحسب گراف هستند.

۸. برای نمونه، رک.

R. C. Tolman, *Principles of Statistical Mechanics*, Oxfotd, Clarendon, 1938.

9. Technical Report No. 65, *The Research Laboratory of Electronics*, M. I. T., March 17, 1949.

پیوست ۱. رشد تعداد دسته‌های نمادها با یک شرط حالت متناهی

فرض کنید $N_i(L)$ تعدادِ دسته‌های به طول L از نمادها باشد که به حالت i منتهی می‌شوند. در این صورت داریم

$$N_j(L) = \sum_{is} N_i(L - b_{ij}^{(s)})$$

که در آن b_{ij}^{1}، b_{ij}^{2}، ...، b_{ij}^{m} طول دسته نمادهایی هستند که ممکن است در حالت i انتخاب شوند و به حالت j منجر شوند. رابطه‌های بالا معادله‌های تفاضلیِ خطی هستند و وقتی $L \to \infty$ باید داشته باشیم

$$N_j = A_j W^L$$

با جانشانی در معادلهٔ تفاضلی داریم

$$A_j W^L = \sum_{i,S} A_i W^{L-b_{ij}^{(s)}}$$

یا

$$A_j = \sum_{iS} A_i W^{-b_{ij}^{(s)}}$$
$$\sum_i \Big(\sum_S W^{-b_{ij}^{(s)}} - \delta_{ij}\Big) A_i = 0$$

برای برقراری این رابطه، دترمینان

$$D(W) = |a_{ij}| = \Big|\sum_S W^{-b_{ij}^{(s)}} - \delta_{ij}\Big|$$

باید صفر باشد و از اینجا W معین می‌شود که، البته، بزرگترین ریشهٔ حقیقی $D = 0$ است.

پس کمیت C به‌صورت

$$C = \lim_{L\to\infty} \frac{\log \sum A_j W^L}{L} = \log W$$

مشخص می‌شود؛ خاطر نشان می‌کنیم که اگر هم همهٔ دسته‌های نمادها در حالتِ (دلخواهِ) واحدی آغاز شوند، همین حکم در بارهٔ ویژگیهای رشد آنها صادق است.

پیوست ۲. استنتاج $H = -\sum p_i \log p_i$

فرض کنید $H(1/n, 1/n, \dots, 1/n) = A(n)$. بنا به شرط (۳) می‌توانیم یک انتخاب از میان s^m امکانِ هم‌احتمال را به m انتخاب از میان s امکانِ هم‌احتمال تجزیه کنیم و به‌دست آوریم

$$A(s^m) = mA(s)$$

همین‌طور

$$A(t^n) = nA(t)$$

می‌توانیم n را به‌قدر دلخواهْ بزرگ اختیار کنیم و mای بیابیم که در

$$s^m \le t^n < s^{(m+1)}$$

صدق کند. پس با گرفتن لگاریتم و تقسیم بر $n \log s$ داریم

$$\left|\frac{m}{n} - \frac{\log t}{\log s}\right| < \epsilon \quad \text{یا} \quad \frac{m}{n} \le \frac{log\, t}{\log s} \le \frac{m}{n} + \frac{1}{n}$$

که در آن ϵ به قدر دلخواهْ کوچک است. حال بنا به‌ویژگیِ یکنواییِ $A(n)$،

$$A(s^m) \le A(t^n) \le A(s^{m+1})$$
$$mA(s) \le nA(t) \le (m+1)A(s)$$

پس با تقسیم بر $nA(s)$ خواهیم داشت

$$\left|\frac{m}{n} - \frac{A(t)}{A(s)}\right| < \epsilon \quad \text{یا} \quad \frac{m}{n} \le \frac{A(t)}{A(s)} \le \frac{m}{n} + \frac{1}{n}$$
$$\left|\frac{A(t)}{A(s)} - \frac{\log t}{\log s}\right| \le 2\epsilon \qquad A(t) = -K\log t$$

که در آن K باید عدد مثبتی باشد تا در (۲) صدق کند.

اکنون فرض کنید انتخابی از میان n امکان با احتمالهای متوافق $p_i = n_i/\sum n_i$ داریم که در آن n_iها عددهای صحیح‌اند. می‌توانیم یک انتخاب از میان $\sum n_i$ امکان را به انتخابی از n امکان با احتمالهای p_1، ...، p_n و آنگاه، اگر iاُمی انتخاب شد، به انتخابی از n_i امکان با احتمالهای برابر، تجزیه کنیم. باز با استفاده از شرط ۳، انتخاب کلی از $\sum n_i$ را که به‌وسیلهٔ دو روش محاسبه شده با هم مساوی قرار می‌دهیم

$$K \log \sum n_i = H(p_1, \dots, p_n) + K \sum p_i \log n_i$$

پس

$$H = K\big[\sum p_i \log \sum n_i - \sum p_i \log n_i\big]$$
$$= -K \sum p_i \log \frac{n_i}{\sum n_i} = -K \sum p_i \log p_i$$

اگر p_iها نامتوافق باشند، می‌توان آنها را با عددهای گویا تقریب زد و همین فرمول باید بنا به فرض پیوستگی ما برقرار باشد. پس فرمول در حالت کلی درست است. انتخاب ضریب K چنان صورت می‌گیرد که برای منظور ما مناسب باشد و در حکمِ انتخاب واحدی برای شاخص است.

پیوست ۳. قضیه‌هایی در بارهٔ منابع ارگودیک

فرض می‌کنیم منبعْ ارگودیک است و بنابراین، قانون قوی اعداد بزرگ را می‌توان در مورد آن به‌کار برد. پس تعداد دفعاتی که یک مسیر مفروض p_{ij} در شبکه دز دنباله‌ای طولانی به طول N طی می‌شود تقریباً متناسب با احتمالِ بودن در i، مثلاً P_i، و پس از انتخاب این مسیر، تقریباً متناسب با $P_i p_{ij} N$ است. اگر N به‌قدر کافی بزرگ باشد، احتمال خطا برحسب درصد، $\pm\delta$، در اینجا کمتر از ϵ است و در نتیجه، بجز در موردِ مجموعه‌ای با احتمالِ کم، همواره عددها عملاً در محدودهٔ

$$(P_i p_{ij} \pm \delta)N$$

قرار می‌گیرند. پس تقریباً همهٔ دنباله‌ها احتمال pای دارند که برابر است با

$$p = \Pi p_{ij}^{(P_i p_{ij} \pm \delta)N}$$

و $\log p/N$ به‌صورت زیر محدود می‌شود

$$\frac{\log p}{N} = \sum (P_i p_{ij} \pm \delta) \log p_{ij}$$

یا

$$\left| \frac{\log p}{N} - \sum P_i p_{ij} \log p_{ij} \right| < \eta$$

قضیهٔ ۳ به این ترتیب به اثبات می‌رسد.

قضیهٔ ۴ از اینجا با محاسبهٔ کرانهای بالا و پایین $n(q)$ براساس دامنهٔ مقادیر ممکن p در قضیهٔ ۳ مستقیماً نتیجه می‌شود: در حالت مختلط (غیرارگودیک) اگر

$$L = \sum p_i L_i$$

و آنتروپی‌های مؤلفه‌ها $H_1 \geq H_2 \geq \ldots \geq H_n$ باشند، داریم

قضیه. *تابع* $\lim_{N\to\infty} \log n(q)/N = \varphi(q)$ *یک تابع پله‌ای نزولی است، $\varphi(q)$ در بازهٔ $\sum_1^{s-1} \alpha_i < q < \sum_1^s \alpha_i$ برابر است با H_s.*

برای اثبات قضیه‌های ۵ و ۶ نخست توجه کنید که F_N یکنوایِ نزولی است زیرا با افزایش N، یک اندیس به یک آنتروپی شرطی اضافه می‌شود. با جانشانی ساده‌ای برای $p_{B_i}(S_j)$ در تعریف F_N نشان داده می‌شود که

$$F_N = NG_N - (N-1)G_{N-1}$$

و با جمع‌کردنِ این برابریها برای همهٔ Nها، به‌دست می‌آوریم

$$G_N = \frac{1}{N} \sum F_N$$

پس $G_N \geq F_N$ و G_N یکنوایِ نزولی است. همچنین آنها باید به حدِ واحدی میل کنند. با استفاده از قضیهٔ ۳ می‌بینیم که $\lim_{N\to\infty} G_N = H$