بسم الله الرحمن الرحيم

دانشکده علوم ریاضی و مهندسی کامپیوتر

پایان‌نامه برای اخذ درجه کارشناسی ارشد آمار
(گرایش ریاضی)

عنوان:

نرخ آنتروپی در فرآیندهای الگویی

استاد راهنما:

دکتر عین‌اله پاشا

تدوین:

مرتضی رئیسی

شهریور ۱۳۸۷

تقدیم به:

روح پدرم که آرزوی دیدن چنین روزی را داشت،

مادر فداکارم که هر چه دارم از اوست،

و آئینه زندگی‌ام.

تقدیر و تشکر

سپاس خداوند منان را که زندگی کردن به من آموخت. حقیقت امر این است که ما کاره‌ای نیستیم. بدین نکته معترف نبودن خامی و پوچی بسیار می‌خواهد. و پرسشی پیش می‌آید که: پس چه می‌گویی؟

برای این پرسش پاسخی اندیشیده‌ایم. از هر چه بگذریم بالاخره ما هم تماشایی این زندگی و زمانه‌ایم.

از اساتید محترم، آقای دکتر عین‌اله پاشا که در طول این دوره و تدوین این پایان‌نامه از راهنمایی‌های ایشان بهره‌ها بردم و همچنین آقای دکتر علی‌اکبر رحیم‌زاده ثانی که زحمات زیادی در طی این دوره برایم کشیدند، کمال تشکر را دارم.

برای آقای دکتر شهرام منصوری که داوری این پایان‌نامه را بر عهده داشتند، توفیق روزافزون از ایزد منان خواستارم.

از تمامی دوستانم که در تدوین این پایان‌نامه مرا یاری کردند از جمله آقایان: علی‌رضا چاجی، حمیدرضا طاهری‌زاده زارچ، حمزه ابراهیمی، مرتضی آقابابایی جزی، مجید راهرو زرگر، مصطفی داوطلب علیایی، علی جباری، احسان قاسمی و ... تشکر و قدردانی می‌کنم.

چکیده

در این پایان‌نامه نرخ آنتروپی فرآیندهای تصادفی و فرآیندهای الگویی القایی را مطالعه کرده و ارتباط بین نرخ آنتروپی این دو را مشخص می‌کنیم. این ارتباط را در حالتی که فرآیند تصادفی اصلی، فرآیند ارگودیک مانا، مارکوف و نویزدار باشد با محدودیت‌های مشخص روی الفباهای القا شده، بررسی می‌شود. همچنین برای حالتی که نرخ آنتروپی فرآیند الگویی نامتناهی باشد، نرخ رشد آنتروپی فرآیند الگویی را مشخص می‌کنیم و نشان می‌دهیم که نرخ آنتروپی فرآیند الگویی القایی از فرآیندی که بصورت $\widehat{X}_i \sim f^*$ (مستقل و همتوزیع) توزیع شده باشد با آنتروپی توزیع f^* برابر است. در پایان مساله بازی را مطرح می‌کنیم.

واژگان کلیدی: نرخ آنتروپی، فرآیند الگویی، الفبا، نرخ رشد، آنتروپی.

فهرست مطالب

پیش گفتار

امروزه آنتروپی و نظریه اطلاع در رشته‌های آمار، مهندسی مکانیک، برق، نجوم، اقتصاد، پزشکی، مدیریت و ... کاربرد دارد. آنتروپی از واژه یونانی Entropy به معنی "به درون خود می‌روم" گرفته شده است، این اصطلاح اولین بار توسط کلوزیوس[۱] [۲۲] در ترمودینامیک به کار گرفته شد.

در سال ۱۹٤۸ نظریه اطلاع امروزی توسط شانون[۲] پایه‌گذاری شد. شانون در [۱۷] نرخ آنتروپی فرآیندهای تصادفی را معرفی کرد. در واقع آنتروپی یک متغیر تصادفی، تعداد متوسط بیت‌های لازم از اطلاعات برای توصیف متغیر تصادفی مذکور می‌باشد و نرخ آنتروپی فرآیند تصادفی، تعداد متوسط بیت‌های لازم از اطلاعات که هر نشانه بتواند فرآیند را توصیف کند.

پایان‌نامه‌ای که پیش رو دارید، شامل سه فصل است که در فصل اول مقدماتی از آنتروپی و فرآیندهای تصادفی و نرخ آنتروپی فرآیندهای تصادفی را بیان می‌کنیم.

در فصل دوم فرآیند الگویی را معرفی می‌کنیم و توزیع احتمال آن را بدست می‌آوریم.

در فصل سوم که شالوده اصلی این پایان‌نامه می‌باشد، نرخ آنتروپی فرآیندهای الگویی القایی فرآیندهای تصادفی را در حالت‌هایی که فرآیند تصادفی، فرآیندی مارکوف، ارگودیک مانا و نویزدار باشد با محدودیت‌های مشخص روی الفباهای القا شده بدست می‌آوریم و ارتباط بین نرخ آنتروپی فرآیند تصادفی و فرآیند الگویی القایی آن را بررسی می‌کنیم.

[۱] Clausius
[۲] Shannon

شمیر[۱] اولین بار ارتباط بین نرخ آنتروپی فرآیندهای تصادفی و فرآیندهای الگویی القایی آن را در [۱۶] برای فرآیندهای i.i.d بررسی کرد. شمیر کران‌هایی برای آنتروپی فرآیندهای الگویی بدست آورد که طبیعتا می توان این کران‌ها را برای نرخ آنتروپی بکار برد. این کران‌ها در [۱۵],[۱۴],[۱۳] اصلاح شدند. در[۱۶],[۱۵],[۱۴],[۱۳] آنتروپی فرآیند الگویی متناهی فرض می‌شود ولی در مقاله‌های مذکور بطور کامل ارتباط بین نرخ آنتروپی فرآیند تصادفی و فرآیند الگویی القایی آن مشخص نشده است. در این پایان‌نامه این ارتباط را در حالت‌های کلی‌تر (مارکوف، ارگودیک مانا و نویزدار) مورد بررسی قرار می‌دهیم.

این پایان‌نامه بر اساس مقالات زیر تدوین شده است:

[1] G. M. Gemelos and T. Weissman, On the Entropy Rate of Pattern Processes, Hewlett Packard Tech Report HPL-2004-159, September 2004.

[2] A. Orlitsky, N. P. Santhanam, and J. Zhang, Universal compression of memoryless source over unknown alphabets, IEEE Transactions on Information Theory, IT-50:7 (July 2004), p. 1469 – 1481.

[۱] Shamir

فصل اول

کلیات

۱-۱ مقدمه

پایه و اساس مفهوم آنتروپی در سال ۱۸۷۷ توسط لئودوینگ بولتزمان[1] مطرح شد [۳] و بعدها در سال ۱۹٤۸ بوسیله‌ی کلود شانون[2] و قبل از وی افراد دیگری چون اچ. نیکویست[3] در تفسیر آنتروپی از نقطه نظر تئوری اطلاع تلاش‌هایی چند انجام شد. نیکویست در سال ۱۹۲٤ 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 دگر شکلی فراهم نمود [۳].

ار. وی. ال. هارتلی[4] اولین کسی بود که کوشید اندازه اطلاع را تعریف کند. هارتلی مقدار اطلاع را به صورت لگاریتم تعداد پیام‌های قابل تشخیص تعریف می‌کند. بنابراین در حالتی که پیام‌ها با طول k باشند، داریم:

$$H_H(s^k) = \log(s^k) = k\log(s).$$

مشکل اساسی رهیافت هارتلی این بود که در آن هیچ فرضی در مورد امکان رخ دادن s نماد با احتمال‌های نابرابر یا امکان وجود وابستگی بین نمادهای متوالی نشده است. دستاورد شانون این است که او نظریه‌های نیکویست و هارتلی را توسعه داد و نظریه اطلاع امروزی را با مرتبط ساختن اطلاع با عدم حتمیت، با بهره‌وری از مفهوم شانس یا احتمال پایه‌گذاری کرد.

[1] -Ludwing Boltzman
[2] -Claude Shannon
[3] -H.Nyquist
[4] -R.V.L.Hartley

نظریه اطلاع که امروزه از آن صحبت می‌شود، برای نخستین بار توسط کلود شانون در سال ۱۹٤۸ مطرح شد. انتشار نظریه‌ی آنتروپی در نظریه اطلاع توسط شانون مؤید این عقیده بود که " بعضی از متغیرهای تصادفی، تصادفی‌تر از متغیرهای دیگرند ". آنتروپی میزان دوری یک متغیر تصادفی از یک مقدار خاص را به صورت کمی در می‌آورد. در واقع می‌توان گفت که مفهوم آنتروپی شانون هسته اصلی نظریه اطلاع را تشکیل می‌دهد. آنتروپی از واژه یونانی Entropy به معنی "به درون خود می‌روم" گرفته شده است. این اصطلاح اولین بار توسط کلوزیوس به کار برده شد. آنتروپی را می‌توان با بی‌نظمی معادل دانست، هرچه نظم سیستم بالا رود آنتروپی آن کاهش می‌یابد و بالعکس.

برای نشان دادن اطلاع به صورت کمی، نظریه اطلاع به وجود آمد. میزان اطلاعی که هر موضوع به ما می‌دهد به وسیله نسبت تعداد سوالات لازم برای پی‌بردن به آن موضوع، اندازه گیری می‌شود.

بعنوان مثال اگر موضوعی که مورد نظر است، در فضایی غیر هم‌شانس قرار داشته باشد متوسط تعداد سؤال‌هایی را که برای رسیدن به موضوع لازم است اطلاع شانون (آنتروپی شانون) گوییم و با $H(X)$ یا $H(f)$ نشان می‌دهیم. به صورت شهودی مقدار اطلاعات دریافتی از وقوع یک حادثه با احتمال وقوع این حادثه نسبت عکس دارد. بعبارت دیگر پیام مربوط به حادثه‌ای که دارای کمترین احتمال وقوع است، بیشترین اطلاع را دارا است. اغلب آنتروپی را میزان تعجب یا میزان عدم حتمیت می‌نامند.

۱-۲ میزان تعجب، عدم حتمیت و آنتروپی

پیشامد E را در نظر می‌گیریم که می‌تواند رخ دهد، اگر آزمایش انجام گیرد و گفته شود که E رخ داده است چقدر تعجب می‌کنیم؟

برای مثال، اتفاق شگفت‌انگیزی که در سال ۱۹۳۸ روی داد صید یک ماهی بود که به گمان دانشمندان سالیان قبل از آن منقرض شده بود[۱]. صید این ماهی باعث به وجود آمدن تغییر اساسی در نظریه تکامل در زیست شناسی شد. در حالی که صید یک شاه ماهی که غیر ممکن نیز به نظر نمی‌رسد به سختی می‌تواند باعث ایجاد تغییر در نظریه‌ها شود.

پس به طور کلی رخ دادن پیشامدی که حتمیت کمتری دارد، باعث تعجب بیشتر ما خواهد شد، یعنی میزان تعجب تابعی از احتمال است.

حال سعی می‌کنیم میزان تعجب را کمی کنیم در واقع ضابطه‌ای برای میزان تعجب بدست می‌آوریم. میزان تعجب را با $S(p)$ نمایش می‌دهیم، یعنی میزان تعجبی که از وقوع پیشامدی با احتمال p حاصل می‌شود و بر این اصل توافق می‌کنیم که میزان تعجب حاصل از رخ دادن E تنها به احتمال وقوع آن بستگی دارد. فرض می کنیم $S(p)$ برای همه مقادیر $۰ < p \leq ۱$ تعریف شده و برای پیشامدهایی که $p = ۰$ است، تعریف نشده است. در این قسمت سعی می‌کنیم که شکل تابعی $S(p)$ را براساس توافق روی مجموعه‌ای از شرایط منطقی تعیین کنیم و ثابت کنیم که این اصول برای شکل معین $S(p)$ لازم هستند.

---

[۱] -Coelacanth

اصل ۱. $S(1)=0$.

اصل ۲. $S(p)$ تابعی اکیداً نزولی از p است، یعنی اگر $p<q$ باشد، آنگاه $S(p)>S(q)$.

اصل ۳. $S(p)$ تابعی پیوسته از p است. در توضیح این شرط این که، انتظار می‌رود هر تغییر کوچک در p باعث تغییر کوچکی در $S(p)$ می‌شود.

اصل ٤. به ازای $0<p\leq 1$ و $0<q\leq 1$، $S(pq)=S(p)+S(q)$. در توضیح این شرط این که، دو پیشامد مستقل E و F را در نظر می‌گیریم به طوری که $P(E)=p$ و $P(F)=q$. پس احتمال وقوع توام برابر $P(EF)=pq$ است. در نتیجه $S(pq)$ میزان تعجب حاصل از وقوع توام E و F است. حال فرض کنیم ابتدا پیشامد E رخ داده و پس از آن پیشامد F رخ داده است. حال چون $S(p)$ میزان تعجب وقوع پیشامد E است، بنابراین $S(pq)$-$S(p)$ نشان دهنده افزایش تعجب است وقتی مطلع می‌شویم که F رخ داده است. بعلاوه چون E و F مستقل از هم هستند، بنابراین افزایش تعجب بایستی برابر با $S(p)$ باشد.

حال قضیه زیر را که بیان کننده ضابطه تابع $S(p)$ است، مطرح می‌کنیم:

قضیه ۱-۱. تابع $S(p)$ در اصول ۱ تا ٤ صدق می‌کند، اگر و تنها اگر $S(p)=-c\log p$ که c یک عدد دلخواه مثبت است و پایه لگاریتم دلخواه و بزرگتر از یک است.

برهان. ر.ج. [۲۰]

برای سادگی محاسبات c را برابر یک قرار می‌دهیم.

عدم حتمیت، سختی پیش بینی برآمدهای متغیر تصادفی X را تشریح می‌کند. فرض کنید متغیر تصادفی X مقادیر $x_1, x_2, \ldots, x_n$ را با احتمال‌های متناظر $p_1, p_2, \ldots, p_n$ اختیار کند. ما به دنبال یافتن کمیتی هستیم که عدم حتمیت X را اندازه‌گیری نماید. عدم حتمیت را که تابعی از $p_1, p_2, \ldots, p_n$ است با $H(p_1, p_2, \ldots, p_n)$ یا به اختصار با $H(X)$ نشان می‌دهیم. $H(p_1, p_2, \ldots, p_n)$ را می‌توان به صورت میانگین وزنی اعداد $S(p_1), S(p_2), \ldots, S(p_n)$ در نظر گرفت که $S(p_i)$ عدم حتمیت پیشامد $\{X = x_i\}$ یا بعبارت دیگر میزان تعجب حاصل از اطلاع وقوع پیشامد $\{X = x_i\}$ است. تابع $H(X)$ متوسط عدم حتمیت پیشامدهای $\{X = x_i\}$ است.

از آنجایی که $-\log p_i$ نشان دهنده میزان تعجب حاصل از این است که متغیر تصادفی X مقدار x_i را اختیار کند، در نتیجه میانگین وزنی میزان تعجب حاصل از اطلاع در مورد متغیر تصادفی X برابر است با

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i .$$

$H(X)$ در نظریه اطلاع به عنوان آنتروپی متغیر تصادفی X شناخته می‌شود. در صورتی که یکی از احتمالات برابر صفر باشد، باز هم مشکلی پیش نمی‌آید زیرا $\lim_{u \to 0} u \log u = 0$.

تعریف ۱-۱. فرض کنید X یک متغیر تصادفی گسسته با تکیه‌گاه S_X و تابع جرم احتمال $f(x)$ باشد، در این صورت آنتروپی X به صورت زیر تعریف می‌شود:

$$H(X) = -E\left(\log f(X)\right)$$

$$= -\sum_{x \in S_X} f(x) \log f(x).$$

یادآوری این نکته ضروری است که آنتروپی X تابعی از احتمالات $f(x_i) = P(X = x_i)$ است و به مقادیر X بستگی ندارد، لذا به جای نماد $H(X)$ می‌توان از نماد $H\big(f(X)\big)$ یا $H(f)$ یا $H(p_۱, p_۲, \ldots)$ استفاده کرد.

در تعریف ۱-۱ پایه لگاریتم تعیین نشده است. در اکثر موارد مهم نیست که آن را چه عددی قرار دهیم، چرا که تغییر در پایه لگاریتم صرفاً تغییر در مقیاس واحدها می‌باشد. اما اغلب پایه را ۲ یا e (عدد نپر) قرار می‌دهند. واحد آنتروپی را هنگامی که در پایه ۲ محاسبه می‌شود bit می‌نامند و هنگامی که در پایه e محاسبه شود، واحد اطلاع بر حسب nat (واحد طبیعی) بیان می‌شود. رابطه بین bit و nat به صورت زیر است:

$$۱ \text{ bit} = ۰/۶۹۳ \text{ nat}.$$

مثال ۱-۱. آنتروپی متغیر تصادفی برنولی X با پارامتر p به صورت زیر است:

$$H(X) = -\sum_{x=۰}^{۱} p^x (۱-p)^{۱-x} \log(p^x (۱-p)^{۱-x})$$

$$= -(۱-p) \log (۱-p) - p \log p$$

$$= \log \frac{(۱-p)^{(۱-p)}}{p^p}.$$

به صورت شهودی $H(f)$ عبارت است از عدم قطعیت موجود در f برای قابل پیش بینی بودن برآمدی از X. آنتروپی یکنواختی یک توزیع را اندازه می‌گیرد. با افزایش $H(f)$، $f(x)$ به توزیع یکنواخت نزدیک‌تر می‌شود.

۱-۳ خواص آنتروپی

لم ۱-۱. فرض کنیم $p_۱,p_۲,\cdots,p_n$ و $q_۱,q_۲,\cdots,q_n$ اعداد مثبت دلخواهی باشند. به طوری که $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i$ در این صورت:

$$-\sum_{i=1}^{n} p_i \log p_i \le -\sum_{i=1}^{n} p_i \log q_i \ .$$

و تساوی برقرار است اگر و فقط اگر $(i=1,2,...,n)$ $p_i = q_i$.

برهان. ر.ج. [۳]

لم ۱-۲. به ازای هر $z > ۰$ ، داریم: $\ln z \le z - ۱$.

برهان. ابتدا مشتق دوم تابع $f(z) = \ln(z)$ را بدست می‌آوریم:

$$f(z) = \ln(z) \Rightarrow f'(z) = \frac{۱}{z}$$

از آنجا:

$$f''(z) = -\frac{۱}{z^۲} < ۰$$

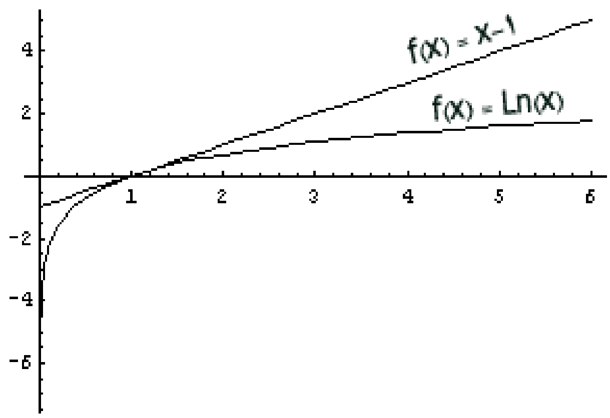
از منفی بودن مشتق دوم بازای تمام $z > ۰$ نتیجه می‌گیریم که $f(z)$ تابع محدب رو به پایین[۱] است. از این رو خط مماس بر منحنی در نقطه‌ی $z = ۱$ در بالای منحنی قرار دارد. از آنجا که خط مماس بر منحنی در نقطه‌ی به طول یک برابر $g(z) = z - ۱$ است (شیب خط برابر است با $f'(۱) = ۱$)

[۱] -convex of downword

خواهیم داشت:

$$f(z) \le g(z) \Rightarrow \ln(z) \le z - ۱$$

و در نتیجه این لم اثبات می‌شود. این مطلب را می‌توان در شکل ۱-۱ ملاحظه کرد.



شکل ۱-۱

قضیه ۱-۲. فرض کنیم $S_X = \{x_۱, x_۲, \ldots, x_n\}$، تکیه‌گاه متغیرتصادفی X باشد و مولفه‌های بردار احتمال $P = (p_۱, p_۲, \ldots, p_n)$ بترتیب احتمالات پیشامدهای $\{X = x_i\}$ باشند، آن گاه:

الف) آنتروپی X کمیتی نامنفی است. یعنی $H(X) \ge ۰$.

ب) آنتروپی X صفر است اگر و تنها اگر X در وضعیت قطعیت کامل باشد.

ج) حداکثر آنتروپی هنگامی است که X دارای توزیع یکنواخت باشد. به عبارت دیگر $H(X) \le \text{Log } n$ و تساوی برقرار است اگر و تنها اگر $p_۱ = p_۲ = \cdots = p_n = \frac{۱}{n}$. همچنین با افزایش اعضای S_X، آنتروپی افزایش می‌یابد.

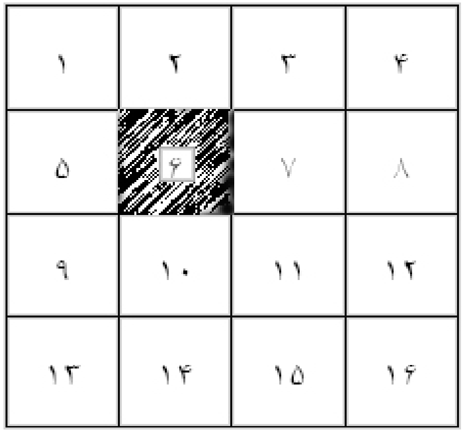
برهان. ر.ج. [۳]

قضیه ۱-۳. $H(P)$ بر حسب P مقعر است.

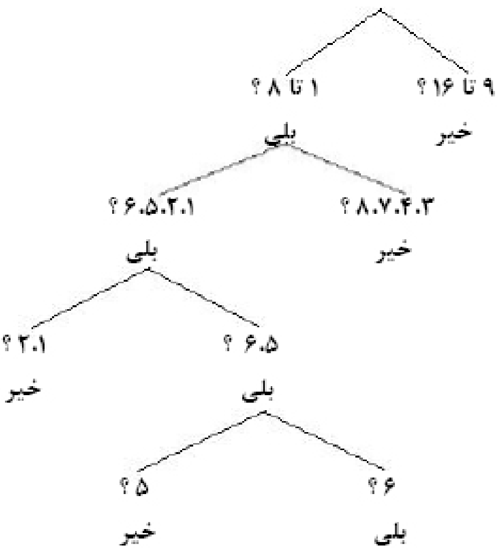
برهان. ر.ج. [۳]

برای درک بیشتر مفهوم آنتروپی شانون به دو مثال زیر توجه کنید:

مثال ۱-۲. فرض کنید منطقه‌ای شامل ۱٦ ناحیه است که یکی از آن ها سایه‌دار است (شکل ۱-۲).



شکل ۱-۲



شکل ۱-۳

با پرسش سؤالاتی که تنها می‌توان با بلی و خیر پاسخ داد، می‌خواهیم تعیین کنیم که ناحیه 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 ناحیه‌ی سایه‌دار را بیابیم ۱۵ 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 به پرسش و پاسخی که در ادامه آورده شده توجه کنید (شکل ۱-۳):

۱- آیا ناحیه سایه‌دار یکی از ۸ ناحیه پایین شکل است؟

پاسخ: "خیر" ، بنابراین ناحیه‌های ۹ تا ۱٦ را می‌توان حذف کرد.

۲- آیا ناحیه سایه‌دار یکی از ٤ ناحیه باقی مانده‌ی سمت چپ است؟

پاسخ: "بلی" ، بنابراین ناحیه سایه‌دار ۱،۲،۵ یا ٦ است.

۳- آیا ناحیه سایه‌دار یکی از دو ناحیه‌ی پایین این ٤ ناحیه باقی مانده است؟

پاسخ: "بلی" ، از این رو ناحیه سایه‌دار ۵ یا ٦ است.

٤- آیا ناحیه سمت چپ است؟

پاسخ: "خیر" ، بنابراین ناحیه سایه‌دار ٦ است.

بنابراین برای تعیین اینکه کدام یک از ۱٦ ناحیه سایه‌دار است، ٤ سؤال لازم است.

اکنون اگر با توجه به این مساله مقدار اطلاع را بررسی کنیم، چون تمام ۱٦ ناحیه دارای احتمال مساویند، داریم:

$$H(X) = -\sum_{i=۱}^{۱۶} \frac{۱}{۱۶} \log \frac{۱}{۱۶} = ۴ \text{ bit},$$

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

این مطلب را در مثال بعد وقتی توزیع احتمال به گونه‌ای است که در آن همه‌ی احتمال‌ها برابر نیستند، بررسی می‌کنیم.