بسم الله الرحمن الرحیم

دانشگاه شیراز

دانشکده علوم

پایان‌نامه کارشناسی ارشد در رشته‌ی آمار

بررسی آنتروپی گذشته تعمیم یافته

توسط:

فاطمه غربی

استاد راهنما:

دکتر عبدالرضا بازرگان لاری

شهریور ماه ۱۳۸۶

به نام خدا

بررسی آنتروپی گذشته تعمیم یافته

به وسیله‌ی :

فاطمه غربی

پایان‌نامه

ارائه‌شده به تحصیلات تکمیلی دانشگاه به‌عنوان بخشی

از فعالیت های لازم برای اخذ درجه کارشناسی ارشد

در رشته‌ی:

آمار ریاضی

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان‌نامه با درجه : عالی

دکتر عبدالرضا بازرگان لاری، استادیار بخش آمار (رئیس کمیته)..........

دکتر جواد بهبودیان،استاد بخش آمار..........

دکتر زهره شیشه بر، استادیار بخش آمار..........

شهریورماه ۱۳۸۶

تقدیم به

پدر و مادر عزیز و همسر مهربانم که همواره پشتیبان و مشوق من بوده اند.

سپاسگزاری

خداوند منان را سپاس گفته که به فضل و لطفش توانسته ام بخش دیگری از راه بی پایان و پرپیچ و خم کسب علم را پشت سر نهاده و گامی در عرصه علم و تحقیق بردارم.در اینجا خود را موظف می دانم که از بزرگوارانی که این راه را برایم هموار ساختند تشکر نمایم.از استاد راهنمای عزیز جناب آقای دکتر عبدالرضا بازرگان لاری که با دقت و حوصله پاسخگوی سوالات این جانب بوده اند ،کمال تشکر وقدردانی را دارم.همچنین از اساتید محترم جناب آقای دکتر جواد بهبودیان و سرکار خانم دکتر زهره شیشه بر که استفاده از راهنمایی های ازرنده ایشان در پربار تر کردن تحقیق بسیار موثر بوده است،سپاسگزارم.از کارکنان بخش آمار،سرکار خانم میرزایی،سرکار خانم ریحانی و جناب آقای شیخ عطار کمال تشکر و قدردانی را دارم و امیدوارم که این عزیزان همواره سلامت بوده و در راه خدمت به اسلام در این آب و خاک علوی موفق و پیروز باشند.

چکیده

بررسی آنتروپی گذشته تعمیم یافته

به وسیله‌ی:

فاطمه غربی

این پایان نامه در مورد آنتروپی گذشته تعمیم یافته وکاربرد آن می باشد.در فصل اول نظریه اطلاع و آنتروپی را مرور کرده و در فصل دوم خصوصیات متغیر تصادفی طول عمر را بر مبنای آنتروپی گذشته تعمیم یافته بیان می کنیم.در فصل سوم یک کلاس جدید از توزیع ها بر اساس آنتروپی گذشته تعمیم یافته معرفی نموده و در فصل چهارم تشخیص یابی برخی توزیع ها را بر مبنای آنتروپی گذشته تعمیم یافته شرح می دهیم.در فصل پنجم مطالب فوق را برای توزیع های گسسته بررسی می کنیم.سرانجام در فصل ششم به ذکر دو مثال می پردازیم.

فهرست مطالب

عنوان .... صفحه

فهرست جدول‌ها

فهرست نمودارها

فصل اول:
مروری بر نظریه اطلاع و آنتروپی

فصل اول:مروری بر نظریه اطلاع و معرفی آنتروپی

۱-۱- مقدمه

نظریه اطلاع علمی است که با مفهوم اندازه و کاربرد اطلاع سر وکار دارد.به طور کلی سه نوع اطلاع وجود دارد:

- اطلاع ترکیبی: علاماتی هستند که پیام ها باآن ها ساخته می شود.
- اطلاع معانی:وابسته به معانی پیام ها و جنبه های معرفی آن می باشد.
- اطلاع عملی:وابسته به کارگیری و اثر پیام هاست.

طبیعت آن ها چنین است: اطلاع ترکیبی اصولا شکل اطلاع را در نظر می گیرد در حالی که اطلاع معانی و عملی وابسته به محتوای اطلاع می باشد.

جملات زیر را در نظر بگیرید:

الف) شخصی با تاکسی به راه آهن آمد.

ب)تاکسی شخصی را به راه آهن آورد.

ت)ترافیک سنگین در بزرگراه $A3$ بین نورنبرگ و مونیخ آلمان وجود دارد.

ث)ترافیک سنگینی در بزرگراه$A3$ در آلمان وجود دارد.

جملات الف و ب به طور ترکیبی متفاوتند با وجود این به صورت معانی و عملی یکی هستندآنها دارای یک معنی و هردو دارای اطلاع برابرند.

جملات ت و ث نه تنها از نظر ترکیبی متفاوتند بلکه همچنین نسبت به معانی یکی نیستند جمله ت اطلاع دقیق تری نسبت به جمله ث دارد.ما به دنبال یک رهیافت نسبتا تکنیکی برای رسیدن به میزان اطلاع هستیم.

ار.وی.ال.هارتلی[1] در سال۱۹۲۸ اولین فردی است که کوشید نظریه اطلاع را تعریف کند. او در این باره به طریق زیر اقدام کرد.

[1] R.V.I.Hartly

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

برای پیام های به طول ۱ داریم:

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

این نتیجه با درک ذهنی اینکه اطلاع هر پیام به طول l، l برابر اطلاع پیامی به طول ۱ است سازگار می باشد.این مطلب حضور موجه لگاریتم را در تعریف هارتلی نیز بیان می کند.

به سادگی می توان نشان داد که تنها تابعی که در معادله $f(s^l) = lf(s)$ صدق می کند بصورت زیر است:

$$f(s) = \log(s)$$

که اندازه هارتلی برای میزان اطلاع حاصل می گردد.

انتخاب مبنای لگاریتم دلخواه است و بیشتر یک موضوع عادی سازی می باشد اگر از لگاریتم طبیعی استفاده شودواحد اطلاع را نت[1] (واحد طبیعی) می نامند.معمولا عدد دو را به عنوان مبنا انتخاب می کنند.در این صورت واحد اطلاع بر حسب بیت (حاصل از دستگاه دودویی دو مقداری) بیان می شود.در حالت انتخاب دو امکانی وقتی یکی از دو امکان رخ دهدمقدار اطلاع در این صورت برابر با ۱ است.با توجه به این که داریم $\ln X = \dfrac{\log_۲ X}{\log_۲ e}$ نتیجه می شود رابطه بین بیت و نت بصورت زیر است:

$$۱\, nat = ۱/۴۴\, bit$$

در رهیافت هارتلی همان طور که در بالا ذکر شدهیچ فرضی در این مورد که امکان دارد s نماد با شانسهای نابرابر رخ دهد یا اینکه ممکن است بستگی بین نمادهای متوالی وجود داشته باشد در نظر گرفته نشده است.دستاورد بزرگ کلود شانون[2] دانشمند آمریکایی،در سال۱۹۴۸ این است که نظریه هارتلی را توسعه داد و نظریه اطلاع امروزی را با مرتبط ساختن اطلاع با عدم اطمینان با بهره گیری از مفهوم احتمال پایه ریزی کرد.شانون راجع به اطلاع هارتلی پیشنهاد کرد به فرض این که همه نمادها با احتمال برابررخ دهند آن را می توان واقعا به عنوان اندازه اطلاع تفسیر کرد. در حالت کلی شانون اندازه اطلاع را بر مبنای مفهوم احتمال معرفی کرد که اندازه هارتلی را به عنوان حالت خاصی شامل می شود.

---
[1] nat
[2] Claude Shannon

۱-۲- شگفتی،عدم اطمینان،آنتروپی:

در حقیقت ارتباط اطلاع و احتمال غیر منطقی نیست.اگرفضای نمونـه ای را کـه درآن همـه پیشنهادها دارای احتمال رخداد برابرنددر نظر بگیرید،عدم اطمینان زیادی درباره این که کـدام پیــــشامد رخ خواهـــد داد وجـــود دارد.یعنـــی وقتـــی یکـــی از ایـــن پیـــشامدها رخ می دهد اطلاع بیشتری فراهم می کند از حالت هایی که در آن فضای نمونه به طریقی ساخته شده که یک پیشامد با احتمال بالایی رخ دهد.اطلاع از طریق عدم اطمینان بـا مفهـوم شـانس پیوند می خورد.

پیشامد E را در نظر بگیرید که می تواند هنگامی که یک آزمایش انجام می شـود رخ دهـد. اگر بشنویم که در واقع E رخ داده است برایمان چقدر شگفت انگیـز اسـت؟ منطقـی بـه نظـر مـــی رســد کـــه فـــرض شـــود میـــزان ایجـــاد شـــگفتی بـــه وســـیله اطـــلاع رســـانی این که E رخ داده است باید وابسته بـه احتمـال E باشـد.بـرای مثـال اگـر آزمایـشی تـشکیل شده باشد از پرتاب یک جفت تاس همگن آنگاه چندان شگفت انگیز نخواهد بـود کـه بـشنویم پیشامد E رخ داده است هنگامی که E ارائه دهنده این باشد که مجموع خالهای دو تـاس زوج باشد (وبنابراین دارای احتمال $\frac{۱}{۲}$ است)در حالی که مطمئنا برایمان شگفت انگیز تر خواهد بود که بشنویم E رخ داده است هنگـامی کـه E پیـشامد ایـن کـه مجمـوع خالهـای دو تـاس ۱۲ باشد(وبنابراین دارای احتمال $\frac{۱}{۳۶}$ است) .

ما به دنبال این هستیم که مفهوم شگفتی را کمی کنیم. موافقـت مـی کنـیم کـه احـساس شگفتی شخص از رخداد یک پیشامد E فقط وابسته به احتمال E است و میزان شگفتی ایجاد شده توسط وقوع یک پیشامد که دارای احتمال p است را با $s(p)$نشان می دهیم.یک صورت تابعی برای $s(p)$ تعیین می کنیم، بدین ترتیب که ابتدا روی شرایط منطقی که $s(p)$باید در آنها صدق کند توافق می کنیم وآنگاه ثابت می کنیم که این اصول برای اینکه $s(p)$به صورت مشخصی باشد ضروری هستند روی هـم رفتـه $s(p)$بـرای تمـام مقـادیر $۰ < p < ۱$ تعریـف می شود اما برای پیشامدهایی که احتمال آنها صفر است یعنی $p = ۰$ تعریف نمی شود.اینـک به معرفی اصول اندازه اطلاع می پردازیم و آنگاه به کمک این اصول قضایا را اثبات می کنیم.

نخستین اصل فقط عبارت است از واقعیتی شهودی که در هنگام شنیدن اینکه یک پیشامد قطعی رخ داده است هیچ شگفتی وجود ندارد.یعنی

اصل ۱:

$$s(\text{۱}) = \text{۰} \qquad (\text{۱-۱})$$

دومین اصل بیان میکند که شگفتی ایجاد شده برای رخداد پیشامدی که محتملتر از پیشامد دیگراست کمتر از شگفتی ایجاد شده برای رخداد آن پیشامد دیگر است.

اصل۲:

$s(p)$ یک تابع اکیدا نزولی از p است یعنی اگر $p < q$ باشد آنگاه $s(p) > s(q)$ است.

سومین اصل یک عبارت ریاضی از واقعیتی است که به گونه ای شهودی انتظار داریم که یک تغییر کوچک در p تغییری کوچک در $s(p)$ بوجود می آورد.

اصل۳:

$s(p)$ تابعی پیوسته از p است.

برای ایجاد انگیزه در موردآخرین اصل دو پیشامد مستقل E و F را در نظر بگیرید که به ترتیب دارای احتمال های $P(E) = p$ و $P(F) = q$ هستند. چون $P(EF) = pq$ است میزان شگفتی ایجاد شده بوسیله آگاهی از اینکه E و F هر دو با هم رخ داده اند عبارت است از $s(pq)$. اینک فرض کنید که گفته شده است که ابتدا E رخ داده است و آنگاه همچنین پس از آن F رخ داده است.چون $s(p)$ میزان شگفتی منبعث از رخداد E است تنیجه می شود که $s(pq) - s(p)$ ارائه دهنده میزان شگفتی اضافی منبعث از وقتی است که آگاه می شویم F نیز همچنین رخ داده است. هر چند هنگامی که F مستقل از E است دانستن اینکه E رخ داده است هیچ تاثیری دراحتمال وقوع F نداردو بنابراین شگفتی اضافی باید فقط برابر با $s(q)$ باشد.این دلیل ارائه آخرین اصل را پیشنهاد می کند.

اصل۴:

برای مقادیر $\text{۰} < p \le \text{۱}$ و $\text{۰} < q \le \text{۱}$، داریم:

$$s(pq) = s(p) + s(q) \qquad (\text{۱-۲})$$

اینک قضیه ۱-۱که منجر به ساختار $s(p)$ میشود معرفی می کنیم.

قضیه۱-۱:

اگر $s(\cdot)$ در اصول ۱ تا۴ صدق کند آنگاه:

$$s(p) = -C \log_{\text{۲}} p$$

که در آن C یک عدد صحیح مثبت دلخواه است[1].

اثبات:

از اصل ۴ نتیجه میشود که :

$$s(p^{2}) = 2s(p) = s(p) + s(p) \qquad (۱-۳)$$

و با استفاده از استقرای ریاضی نتیجه می شود که :

$$s(p^{m}) = ms(p) \qquad (۱-۴)$$

و همچنین

$$s(p) = s(p^{\frac{1}{n}} p^{\frac{1}{n}} \ldots p^{\frac{1}{n}}) = ns(p^{\frac{1}{n}}) \Rightarrow s(p^{\frac{1}{n}}) = \frac{1}{n} s(p) \qquad (۱-۵)$$

بنابراین از معادله های(۱-۴)و(۱-۵) بدست می آوریم که:

$$s(p^{\frac{m}{n}}) = ms(p^{\frac{1}{n}}) = \frac{m}{n} s(p) \qquad (۱-۶)$$

وقتی که x یک عدد گویای مثبت باشد این معادله معادل است با:

$$s(p)^{x} = x\, s(p) \qquad (۱-۷)$$

اما از پیوستگی تابع s (اصل۳) نتیجه می شود که معادله(۱-۷) برای تمام اعـداد حقیقـی غیـر منفی x برقرار است.

اینـک بـرای هـر p کـه $0 < p \leq 1$ باشـد فـرض کنیـد $x = -\log_{2} p$ باشـد آنگـاه $p = (\frac{1}{2})^{x}$ واز معادله(۱-۳) داریم:

$$s(p) = s((\frac{1}{2})^{x}) = xs(\frac{1}{2}) = -C\log_{2} p \qquad (۱-۸)$$

که بنابر اصول ۱و۲ داریم :

$$C = s(\frac{1}{2}) > s(1) \qquad (۱-۹)$$

متداول است که مقدار C را مساوی ۱ اختیار کنند.

یک متغیر تـصادفی X کـه یکـی از مقـادیر $x_{1}, x_{2}, \ldots, x_{n}$ را بـه ترتیـب بـا احتمالهـای $p_{1}, p_{2}, \ldots, p_{n}$ اختیار میکند در نظر بگیرید.چون $-\log p_{i}$ [2] ارائه دهنده میزان شگفتی ایجـاد شده بوسیله هنگامی که X مقدار x_{i} را اختیار کند است نتیجه می شود که بر پایـه آگـاهی از مقدار X امید ریاضی میزان شگفتی که دریافت خواهیم کرد عبارت است از:

[1] مراجعه شود به Reference[2]

[2] برای مطالب باقی مانده بجای $\log_{2} p$ می نویسیم $\log p$

$$H(X)=-\sum_{i=1}^{n} p_i \log p_i \qquad (۱۰-۱)$$

در نظریه اطلاع کمیت $H(X)$ به عنوان آنتروپی[1] متغیر تصادفی X شناخته می شود[2].(در حالتی که یکی از p_i ها مساوی با صفر باشد مقدار $0\log 0$ را مساوی صفر اختیار می کنیم.)

چون $H(X)$ ارائه دهنده میانگین میزان شگفتی دریافتی شخص بر پایه آگاهی از مقدار X است پس می تواند به عنوان ارائه دهنده میزان عدم اطمینان که برای مقدار X وجود داردتعبیرشود.درواقع در نظریه اطلاع $H(X)$ به عنوان میانگین میزان اطلاع دریافتی وقتی X مشاهده شده است تعبیر می شود. بنابراین میانگین میزان شگفتی منبعث از X،عدم اطمینان X یا میانگین میزان اطلاع بدست آمده از X تماما ارائه دهنده مفهوم یکسانی از آنتروپی X می باشند.

در حالت کلی تعریف زیر را داریم:

$$H_b(X)=-\sum_{i=1}^{n} p_i \log_b p_i \qquad (۱۱-۱)$$

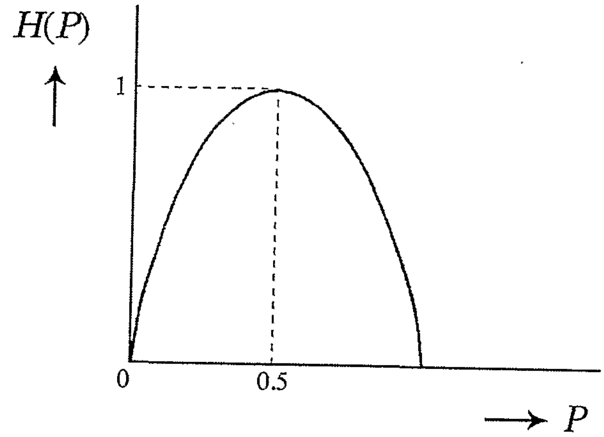
به عنوان یک مثال ساده آزمایش برنولی با دو برآمد به ترتیب با احتمالهای $p_1 = p$ و $p_2 = 1-p$ را در نظر بگیرید آنگاه:

$$H(P)=-p\log p-(1-p)\log(1-p) \qquad (۱۲-۱)$$

نمودار۱-۱ نشان می دهدکه چگونه $H(p)$ به عنوان تابعی از p عمل می کند. می توان نتیجه گرفت که اگر یک برآمد حتمی باشد،یعنی با احتمال ۱ رخ دهد،اندازه اطلاع صفر بدست می آید.این مطلب با درک مستقیم ذهنی که پیشامد های حتمی هیچ اطلاعی را فراهم نمی کنند موافق است.

[1] entropy

[2] $H(P)$ یا $H(X)$ هر دو ارائه دهنده مفهوم یکسانی هستند و بنا به کاربرد هر دو صورت آن را به کار برده ایم.



نمودار ۱-۱ محاسبه آنتروپی توزیع برنولی

قضیه ۱-۲: [1]

فرض کنید $X=(x_۱,x_۲,...,x_n)$ فضای نمونه آزمایش و $P=(p_۱,p_۲,...,p_n)$ توزیع احتمال نظیر آن باشد در این صورت داریم:

$$H(P)\le \log n \qquad (۱۳-۱)$$

تساوی برقرار است اگر وتنها اگر برای تمام مقادیر $i=۱,۲,...,n$ داشته باشیم $p_i=\frac{1}{n}$

و داریم:

$$H(P)\ge ۰ \qquad (۱۴-۱)$$

که تساوی برقرار است اگر وتنها اگر یک k وجود داشته باشد به قسمی که $p_k=۱$ در حالی که برای تمام $i\ne k$، $p_i=۰$ می باشد.

اثبات:

الف)برای اثبات رابطه (۱-۱۳) از نابرابری زیر استفاده خواهد شد:

$$\ln(a)\le a-۱ \qquad (۱۵-۱)$$

برای هر مقدار a که متعلق به مجموعه اعداد حقیقی باشد.

اکنون $H(p)-\log n$ را در نظر بگیرید داریم:

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

$$=-\sum_{i=۱}^{n}p_i\{\log p_i+\log n\}=\sum_{i=۱}^{n}p_i\log(\frac{۱}{p_i n})$$

ازنابرابری $\ln(a)\le a-۱$ نتیجه می شود

[1] Reference[2]

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

که نتیجه می شود

$$H(p)\le \log n$$

تساوی برقرار است اگر وتنها اگر $\frac{1}{p_i n}=1$ که مقدار متناظر با $a=1$ می باشد ،این بدین معنی است که برای تمام مقادیر $i=1,2,\ldots,n$ ، $p_i=\frac{1}{n}$ می باشد.

ب) چون p_i و $-\log p_i$ هردو نمی توانند منفی باشند مقدار اطلاع همواره مثبت یا مساوی صفر است از این رو همواره داریم: $H(p)\ge 0$

به سهولت دیده می شود که $H(p)$ برابر صفر است،اگر یک مولفه p برابر ۱ و بقیه احتمال ها برابر صفر باشند.مقدار ماکزیمم اطلاع در این صورت برابر با $\log n$ است.

□

مثال ۱-۱:

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 می تواند یکی از ۱۰ فاصله شدت را نشان دهد تعداد 10^{414720} تصویر تلویزیونی متفاوت امکان پذیر است.اگر هر یک از این تصاویر با احتمال مساوی رخ دهند،مقدار اطلاع موجود در یک تصویر برابر است با :

$$H(P)=\log n=\log\left(10^{414720}\right)=1/4\times 10^{6}\ \text{bit}$$

مثال ۱-۲:

با این مثال نشان می دهیم که اگر احتمال ها را با ترتیب متفاوتی قرار دهیم ممکن است مقدار اطلاع تغییر نکند.اکنون دو توزیع احتمال زیر را در نظر بگیرید:

$$P=\{0/5,0/25,0/25\} \quad , \quad Q=\{0/48,0/32,0/2\}$$

وقتی مقدار اطلاع را برای هر دو حالت مربوطه حساب کنیم بدست می آوریم:

$$H(P)=H(Q)=1/5bit$$

یعنی توزیعهای احتمال متفاوت می توانند به مقدار اطلاع یکسانی منجر گردند.

۱-۳- آنتروپی و سوالهای بلی-خیر

در این بخش به محاسبه آنتروپی ،در حالتی که با سوال های بلی-خیر روبرو هستیم می پردازیم

مثال ۱-۳:

فرض کنید منطقه ای شامل ۱۶ ناحیه است که یکی از آنها سایه دار است.جدول ۱-۱،با پرسش سوالاتی که تنها می توان با بلی و خیر پاسخ داد،می خواهیم تعیین کنیم که ناحیه سایه دار در کجا قرار دارد. بهترین استراتژی چیست؟می توان حدس زد،ولی در این صورت باید این مخاطره را پذیرفت که قبل از اینکه سرانجام ناحیه سایه دار را بیابیم ۱۶ سوال بپرسیم.بهتر است که به طور انتخابی عمل کنیم.در این صورت بازی پرسش و پاسخ می تواند بصورت زیر پایان پذیرد.برای مثال

۱.آیا ناحیه سایه دار یکی از ۸ ناحیه پایین شکل است؟

پاسخ:خیر،بنابراین ناحیه های ۹تا۱۶ را می توان حذف کرد.

۲.آیا ناحیه سایه دار یکی۴ ناحیه باقی مانده سمت چپ است؟

پاسخ:بلی،بنابراین ناحیه سایه دار ۱ ،۲ ،۵ یا۶ است.

۳.آیا ناحیه سایه دار یکی از دو ناحیه از چهار ناحیه باقی مانده پایین است؟

پاسخ:بلی،از این رو ناحیه سایه دار۵ یا۶ است.

۴.آیا ناحیه سمت چپ است؟

پاسخ،خیر، بنابراین ناحیه سایه دار۶ است.

۱	۲	۳	۴
۵	۶	۷	۸
۹	۱۰	۱۱	۱۲
۱۳	۱۴	۱۵	۱۶

جدول ۱-۱ منطقه ای شامل۱۶ناحیه

در نتیجه برای تعیین این که کدام یک از ۱۶ ناحیه سایه دار است چهار سوال کافی است.

اکنون اگر با توجه به این مساله مقدار اطلاع را بررسی کنیم،چون تمام ۱۶ ناحیه دارای احتمال های مساوی هستند،درمیابیم که:

$$H(X) = -\sum_{i=1}^{16} \frac{1}{16} \log \frac{1}{16} = \log 16 = 4 bit$$