به نام خدای مهدی (عج)

دانشگاه تربیت معلم

دانشکده علوم ریاضی و مهندسی کامپیوتر

پایان‌نامه کارشناسی ارشد آمار

عنوان:

آنتروپی تعمیم یافته گذشته و کاربرد آن در مشخص سازی و مرتب سازی توابع توزیع

استاد راهنما:

جناب آقای دکتر عین‌الله پاشا

تدوین:

مسعود کلهر

شهریور ۱۳۸۶

تقدیم به

پدر صبور

و

مادر مهربان و فدا کارم

تقدیر و تشکر

شکر بیکران خداوند بزرگ و علیم را که توفیق نوشتن این پایان نامه را داد و در واقع تا کنون هر چه که آموخته ام و خواهم آموخت ، قطره ای از اقیانوس بی کران علم و معرفت اوست و امیدوارم که این کار ناچیز برگ رهایی باشد در روز جزا ، برای نیازمند ترین بندگان به رحمت و بخشش خداوندی ، که کسی نیست جز نگارنده .

اعضای خانواده ام مخصوصا پدر و مادرم ، در تمامی مراحل زندگی یار و یاور من بوده اند . از زحمات بی شائبه آنها که فراتر از آن چیزی بود که می توانست باشد ، تشکر و قدر دانی می کنم .

از جناب آقای دکتر عین الله پاشا که در طول این دوره به من کمک های زیادی کردند و زحمات فراوانی متقبل شدند و در واقع ایشان الگوی علمی و اخلاقی برای اینجانب بودند ، نهایت سپاسگزاری را دارم.

در طول این دوره مطالب زیادی از جناب آقای دکتر علی اکبر رحیم زاده ثانی فرا گرفتم و همچنین ایشان زحمت داوری این پایان نامه را بر عهده داشتند که بدینوسیله از ایشان تشکر و قدر دانی می کنم.

همچنین لازم می دانم از جناب آقای دکتر غلامحسین یاری که زحمت داوری این پایان نامه را قبول کردند ،سپاسگزاری کنم.

از تمام دوستانم که در طول این دوره و نگارش پایان نامه به من کمک های زیادی کردند مخصوصا جناب آقایان ، اسرافیل رشدی دیزجیکان ، اسد ا... یوسفی ، احسان قاسمی ، مصطفی داوطلب و حجت رنگین تشکر و قدر دانی می کنم.

چکیده

در متون مربوط به نظریه اطلاع، آنتروپی شانون نقش مهمی را بازی می‌کند. در تحلیل بقا و آزمون طول عمر، معمولاً اطلاعاتی در مورد عمر فعلی مؤلفه‌ی تحت بررسی، در دسترس است. بنابراین سن مؤلفه نیز باید در اندازه‌گیری عدم قطعیت دخیل باشد. با این وجود منطقی است که فرض کنیم عدم قطعیت موقعیت‌های واقعی زیادی لزوماً به آینده مربوط نیستند و می‌توانند به گذشته مربوط باشند. واضح است که در چنین مواردی آنتروپی شانون نمی‌تواند اندازه مناسبی باشد و باید آن را طوری اصلاح نمود که عدم قطعیت به گذشته وابسته باشد. اندازه به‌دست آمده را آنتروپی گذشته نامیده و سپس این آنتروپی گذشته را تعمیم داده و آن را آنتروپی گذشته تعمیم یافته می‌نامیم. بر پایه آنتروپی گذشته تعمیم یافته (آنتروپی گذشته) یک ترتیب آماری تعریف می‌شود. خواص این روش مرتب‌سازی مطالعه شده و ارتباط آن با ترتیب‌های تصادفی کلاسیک مورد بررسی قرار می‌گیرد. همچنین از آنتروپی گذشته تعمیم یافته (آنتروپی گذشته) برای معرفی یک کلاس ناپارامتری استفاده می‌کنیم و خواص کلاس تعریف شده مورد بحث قرار می‌گیرد. در ضمن نشان داده شده است که تحت شرایط خاص، آنتروپی گذشته تعمیم یافته (آنتروپی گذشته) تابع توزیع را به‌طور یکتا مشخص می‌کند.

واژه‌های کلیدی:

اندازه اطلاع، آنتروپی گذشته، آنتروپی مانده، تابع نرخ خطر معکوس شده، IUL ، $IUL(\beta)$ ، روش‌های مرتب سازی LU ، PE و GPE

۱۳۸۷/۲/۱۱

فهرست مطالب

فصل سوم : تعمیم آنتروپی گذشته و کاربردهای آن در تعیین توابع توزیع

مقدمه

از آنجایی که آنتروپی شانون نمی‌تواند اندازه مناسبی برای عدم قطعیت مربوط به گذشته باشد. لذا هدف این پایان‌نامه به‌دست آوردن اندازه‌ای برای سنجش عدم قطعیت گذشته مربوط به متغیرهای تصادفی پیوسته و گسسته است که علاوه بر ویژگی‌های شانون، زمان را نیز در نظر می‌گیرد. سپس تحت شرایط خاصی از این اندازه به‌دست آمده برای مشخص‌سازی توابع توزیع استفاده خواهیم کرد. مطالبی که در این پایان‌نامه ارائه می‌شود به شرح زیر است:

فصل اول: مفاهیم پایه‌ای مبحث آنتروپی

در این فصل ابتدا در بخش نخست تاریخچه‌ای از آنتروپی ارائه کرده و سپس ارتباط میان آنتروپی و نظریه اطلاع را شرح می‌دهیم. در بخش دوم میزان تعجب حاصل از وقوع یک پیشامد را کمی می‌کنیم و با استفاده از یک قضیه مهم اندازه‌ای برای عدم قطعیت متغیرهای تصادفی گسسته به‌دست می‌آوریم. بخش سوم نیز به آنتروپی متغیرهای تصادفی گسسته و تعریف آنتروپی توأم و شرطی اختصاص یافته است که در آن با استفاده از چند قضیه، برخی خواص آنتروپی مورد بحث و بررسی قرار گرفته است. در بخش چهارم با الهام گرفتن از آنتروپی متغیرهای تصادفی گسسته، تعریف آنتروپی شانون را برای متغیرهای تصادفی پیوسته ارائه می‌کنیم و تفاوت‌های بین آنتروپی در حالت‌های گسسته و پیوسته را بررسی می‌کنیم . همچنین روش ماکسیمم آنتروپی را در بخش پنجم معرفی کرده و سپس قضیه تابع چگالی ماکسیمم آنتروپی با قیود گشتاوری را بیان و اثبات می‌نماییم. در بخش آخر نیز تابع اطلاع تشخیص و اطلاع متقابل را تعریف می‌کنیم.

فصل دوم: بعضی از خواص آنتروپی گذشته و کاربردهای آن

در ابتدا آنتروپی مانده را جهت درک روشن‌تر از مفهوم‌های دو نوع آنتروپی، آنتروپی گذشته و آنتروپی مانده، معرفی می‌کنیم. در بخش دوم اندازه‌ای برای سنجش عدم قطعیت گذشته توزیع‌های طول عمر به‌دست می‌آوریم، همچنین تابع نرخ خطر معکوس شده را تعریف خواهیم کرد. در بخش

بعدی از آنتروپی گذشته برای معرفی یک روش جدید مرتب‌سازی (PE) استفاده می‌کنیم که می‌تواند برای مقایسه توزیع‌های طول عمر به‌کار رود. بر پایه آنتروپی گذشته کلاس‌های جدیدی از توزیع طول عمر، یعنی $IUL(DUL)$ و $DRHR(IRHR)$ را در بخش چهارم معرفی کرده و ارتباط آنها را با هم بررسی می‌کنیم. به‌دست آوردن کران‌های بالا و پایین برای آنتروپی گذشته و تابع توزیع قسمت بعدی این بخش را تشکیل می‌دهد. بخش پنجم نیز به مشخص‌سازی توزیع‌های طول عمر با استفاده از آنتروپی گذشته اختصاص یافته و نشان خواهیم داد که اگر آنتروپی گذشته یک متغیر تصادفی پیوسته نامنفی که طول عمر یک مؤلفه را بیان می‌کند، نسبت به زمان نزولی باشد، آنگاه می‌توان به طور یکتا تابع توزیع را مشخص کرد. با الهام گرفتن از آنتروپی گذشته متغیرهای تصادفی پیوسته، آنتروپی گذشته متغیرهای تصادفی گسسته را در بخش ششم معرفی می‌کنیم. سپس با بیان و اثبات یک قضیه مهم توزیع یکنواخت گسسته را به طور یکتا با استفاده از آنتروپی گذشته‌ای که صعودی است، مشخص می‌کنیم. در ادامه بیان می‌کنیم که در حالت کلی آنتروپی گذشته متغیرهای تصادفی گسسته به طور یکتا توزیع احتمال را مشخص نمی‌کند. در انتهای این فصل نیز به مفهوم قابل قبول‌ترین سیستم خواهیم پرداخت.

فصل سوم: تعمیم آنتروپی گذشته و کاربردهای آن در تعیین توابع توزیع

در ابتدای این فصل آنتروپی گذشته تعمیم یافته را معرفی کرده و با استفاده از یک نکته بیان می‌کنیم که آنتروپی گذشته و آنتروپی شانون حالت‌های خاصی از آنتروپی گذشته تعمیم یافته هستند. در بخش بعدی خواص متغیر تصادفی طول عمر بر اساس آنتروپی گذشته تعمیم یافته را مورد بحث و بررسی قرار می‌دهیم. همچنین بر پایه آنتروپی گذشته تعمیم یافته یک روش مرتب‌سازی (GPE) را معرفی خواهیم کرد و نشان می‌دهیم که اگر S' و S دو مجموعه باشند که به ترتیب تمام جفت توزیع‌ها با استفاده از GPE و PE مرتب شده باشند آنگاه $S \subseteq S^*$. همچنین با ارائه چند مثال بررسی خواهیم کرد که هیچ رابطه‌ای میان روش مرتب‌سازی بر اساس آنتروپی گذشته تعمیم یافته و

ترتیب‌های تصادفی کلاسیک وجود ندارند. در بخش سوم به ارائه یک کلاس ناپارامتری از توزیع طول عمر بر اساس آنتروپی گذشته تعمیم یافته، یعنی $(DUL(\beta))IUL(\beta)$ پرداخته و ارتباط این کلاس را با کلاس‌های تعریف شده بر اساس آنتروپی گذشته مورد بررسی قرار خواهیم داد. همچنین در ادامه این بخش نیز یک کران بالا (پایین) برای تابع نرخ خطر معکوس شده برحسب آنتروپی گذشته تعمیم یافته به‌دست می‌آوریم. در بخش‌های چهارم و پنجم نیز بر اساس آنتروپی گذشته تعمیم یافته به مشخص‌سازی توابع توزیع متغیرهای تصادفی پیوسته و گسسته خواهیم پرداخت و نتایجی دقیقاً مشابه با نتایج بخش‌های پنج و شش از فصل دو را به‌دست خواهیم آورد.

این پایان‌نامه بر اساس مقالات زیر انجام شده است:

[1] Nanda, A.K., Paul , P., 2006. Some results on generalized past entropy. Journal of Statistical Planning and Inference 136,3659-3674.

[2] Nanda, A.K., Paul , P., 2006. Some properties of past entropy and their applications. Metrika 64:47-61.

[3] Di Crescenzo , A. , Longobardi , M., 2002. Entropy –based measure of uncertainty in past lifetime distributions. J . Appl . Probab . 39 , 434 – 440.

فصل اول

مفاهیم پایه‌ای مبحث آنتروپی

۱-۱ مقدمه و تاریخچه

نظریه اطلاع برای نخستین بار توسط کلودشانون[۱] در سال ۱۹۴۸ مطرح شد. شانون در سال ۱۹۱۶ در میشیگان ایالات متحده چشم به جهان گشود. وی ۱۵ سال از عمر خویش را در آزمایشگاه بل[۲] صرف کار کردن با افراد مشهوری چون جان پیرس[۳] که به خاطر ارتباطات ماهواره‌ای شهرت فراوانی دارد و هری نیکویست[۴] نمود.

در واقع می‌توان گفت که مفهوم آنتروپی شانون (۱۹۴۸) هسته اصلی نظریه اطلاع را تشکیل می‌دهد که گاهی اوقات تحت عنوان اندازه عدم قطعیت[۵] به آن مراجعه می‌کنیم. این نظریه در آغاز با بسیاری از مسائل ریاضی دشوار روبرو بود که قوانین اولیه شانون دقت کافی برای حل این مسائل را نداشت.

بعدها مک میلان[۶] و فین اشتین[۷] نظریه مفهومات اساسی منابع گسسته (منبع، کد، کانال و ...) را به عنوان اولین تعاریف دقیق ریاضی به کار بردند. کالبک[۸] و لیبلر[۹] در سال ۱۹۵۱ فاصله بین دو چگالی

---

۱) *Claude Shannon*
۲) *Bell Labs*
۳) *John Pierce*
۴) *Harry Nyquist*
۵) *Uncertainty*
۶) *Mc Millan*
۷) *Feinstein*
۸) *Kullback*
۹) *Leibler*

را تعریف کردند که نزدیکی و اطلاع تمییز[1] دو چگالی را نمایش می‌دهد و آنتروپی نسبی[2] نامیده می‌شود، به این ترتیب ارتباط دیگری بین آنتروپی شانون و تئوری استنباط آماری پیدا شد.

جینس[3] در سال ۱۹۵۷ اصل آنتروپی ماکسیمم[4] را بیان نمود. این اصل چگالی آنتروپی ماکسیمم را با قیود گشتاوری داده شده مشخص می‌کرد و به کاربردهای زیادی در علوم و زمینه‌های مختلف از جمله مکانیک آماری، نجوم، هواشناسی، اقتصاد، جغرافیا، تجارت بین المللی، بانکداری، توزیع جمعیتی، پزشکی و ... دست یافت.

کالبک (۱۹۵۹) تابع اطلاع تمییز بین دو چگالی را با استفاده از اصل ماکسیمم آنتروپی بیان نمود که نزدیکی چگالی نامعلوم و چگالی آنتروپی ماکسیمم را نشان می‌دهد.

آنتروپی در کتاب‌های آماری به ندرت مورد بحث و بررسی قرار گرفته است. آنتروپی موضوعی است که به نوعی عدم قطعیت و اطلاع مربوط بوده و در رابطه با ترمودینامیک آماری و رمزگذاری به کار می‌رود.

آنتروپی از واژه یونانی *Entropy* به معنی ((به درون خود می‌روم)) گرفته شده است. این اصطلاح اولین بار توسط کلوزیوس به کار گرفته شد. آنتروپی را می‌توان با بی‌نظمی معادل دانست، هرچه نظم سیستمی بالا رود آنتروپی آن کاهش می‌یابد و بالعکس کاهش نظم باعث افزایش آنتروپی می‌شود.

در ترمودینامیک، آنتروپی اندازه بی‌نظمی در یک سیستم می‌باشد. سیستم از وضعیت آنتروپی پایین به سمت وضعیت آنتروپی بالا حرکت می‌کند، به‌طوری که انرژی غیر قابل استفاده افزایش می‌یابد.

در نظریه اطلاع، آنتروپی اندازه تعداد بایت‌های لازم برای انتقال یک پیام می‌باشد. که این برای سیگنال‌های صوتی روی خط دیجیتال و چیزهای دیگر مفید و کاربردی است. آنتروپی یک سیستم با میزان اطلاع موجود در آن مرتبط است. یک سیستم با نظم بیشتر می‌توانند با بایت‌های کمتری از

---

۱) *Discrimination information*
۲) *Relative entropy*
۳) *Jaynes*
۴) *Maximum entropy*

اطلاعات توصیف شود، در حالی که سیستمی با بی‌نظمی بالاتر برای توصیف شدن به بایت‌های بیشتری از اطلاعات نیازمند است.

نظریه اطلاع در ابتدا برای بیان عددی اطلاع به‌وجود آمد. همان‌طور که فاصله، دما، زمان، و ... را با عدد اندازه‌گیری می‌کنند، مقدار اطلاع را که این موضوع به ما می‌دهد، به‌وسیله تعداد سؤالات مثبت لازم برای پی بردن به آن موضوع اندازه‌گیری می‌شود.

این جواب‌های به‌دست آمده را که به صورت بله و خیر می‌باشد، می‌توان با اعداد صفر و یک نشان داد، به همین دلیل است که واحد اطلاع را بیت (bit[۱]) می‌نامیم.

اگر موضوعی که مورد نظر است در فضایی غیر هم شانس قرار داشته باشد، متوسط تعداد سؤال‌هایی که می‌توان به موضوع مورد نظر رسید را اطلاع شانون (آنتروپی شانون) می‌نامیم و با $H(X)$ نشان می‌دهیم. $H(X)$ بعداً تعریف خواهد شد.

برای به‌دست آوردن یک درک خوب از آنتروپی به عنوان یک اندازه برای کاهش بایت‌های لازم برای مخابره یک پیام مثال زیر را درنظر می‌گیریم.

۱-۱-۱ **مثال.** فرض کنیم مقدار بردار گسسته تصادفی X در موقعیت فرستنده A مشاهده شده و توسط یک شبکه مخابراتی که شامل دو سیگنال صفر و یک است، به موقعیت گیرنده B انتقال داده می‌شود. برای این منظور، ابتدا لازم است که هر مقدار ممکن X را به دنباله‌هایی از صفر و یک تبدیل نموده و یا کدگذاری کنیم. برای جلوگیری از هرگونه ابهامی، لازم است که هیچ دنباله کدگذاری شده را نتوان با استفاده از دنباله کوتاهتری به‌وسیله اضافه نمودن اجزا بیشتر به‌دست آورد.

برای مثال اگر X بتواند مقادیر x_1, x_2, x_3, x_4 را اختیار کند، یکی از روش‌های کدگذاری عبارت است از

۱) *Binary digit*

$$x_1 \leftrightarrow ۰۰$$
$$x_2 \leftrightarrow ۰۱$$
$$x_3 \leftrightarrow ۱۰$$
$$x_4 \leftrightarrow ۱۱$$

یعنی اگر $X = x_1$ ، آنگاه پیام ۰۰ به موقعیت گیرنده B ارسال می‌گردد و اگر $X = x_2$ باشد، پیام ۰۱ به موقعیت گیرنده B فرستاده می‌شود و ... .

حال احتمال‌های زیر را در نظر می‌گیریم

$$P\{X = x_1\} = \frac{1}{2}$$
$$P\{X = x_2\} = \frac{1}{4}$$
$$P\{X = x_3\} = P\{X = x_4\} = \frac{1}{8}$$

آنگاه برای کد داده شده توسط فوق انتظار است به‌طور متوسط بـه ازای هـر مقـدار ممکـن X ، دو بیت ارسال شود $\left[(0/5\times 2+0/25\times 2+0/125\times 2+0/125\times 2)bit = 2bit\right]$.

اما با استفاده از ویژگی مفید نظریه اطلاع می‌توان یک کدگذاری بهتر را یافـت کـه بـه بایـت‌هـای کمتری نیاز داشته باشد. به عنوان مثال، کدگذاری به صورت زیر را در نظر می‌گیریم

$$\begin{array}{ll} x_1 \leftrightarrow ۰ & 0/5\times 1bit = 0/5bit \\ x_2 \leftrightarrow ۱۰ & 0/25\times 2bit = 0/5bit \\ x_3 \leftrightarrow ۱۱۰ & 0/125\times 3bit = 0/375bit \\ x_4 \leftrightarrow ۱۱۱ & 0/125\times 3bit = 0/375bit \end{array}\Bigg\}$$

به طور متوسط $1/75\,bit$ به ازای هر مقدار X

علاوه بر این نظریه اطلاع به ما می‌گوید که آنتروپی این منبع اطلاعات $1/75\,bit$ به ازای هر مقدار X می‌باشد. بنابراین این کدگذاری نسبت به سایر کدگذاری‌ها برای توصیف شدن این منبع اطلاعـاتی بهتر خواهد بود.

۱-۲ تعجب[1] ، عدم قطعیت و آنتروپی

پیشامد E را درنظر می‌گیریم که می‌تواند رخ دهد، یعنی احتمال وقوع آن صفر نباشد، اگر آزمایش انجام گیرد، چقدر تعجب می‌کنیم اگر بشنویم E اتفاق افتاده است؟ به نظر منطقی می‌رسد که تصور کنیم میزان تعجب حاصل از وقوع E تنها به احتمال وقوع E بستگی داشته باشد. برای مثال اگر آزمایش شامل پرتاب یک جفت تاس باشد، اگر بشنویم که مجموع اعداد ظاهر شده زوج است زیاد متعجب نمی‌شویم، زیرا که احتمال وقوع آن برابر $\frac{۱}{۲}$ است، در حالی که اگر بشنویم مجموع اعداد ظاهر شده ۱۲ است بیشتر تعجب می‌کنیم، چرا که احتمال وقوع آن برابر $\frac{۱}{۳۶}$ است.

در این بخش سعی می‌کنیم که میزان تعجب را کمی کنیم.در آغاز روی این اصل توافق می‌کنیم که میزان تعجب حاصل از وقوع E تنها بستگی به احتمال وقوع آن دارد و آن را با $S(p)$ نمایش می‌دهیم، یعنی میزان تعجبی که از وقوع پیشامدی با احتمال p حاصل می‌گردد. حال سعی می‌کنیم که شکل تابعی $S(p)$ را بر اساس توافق روی مجموعه‌ای از شرایط منطقی تعیین نموده و سپس ثابت کنیم که این اصول برای این که $S(p)$ شکل معینی داشته باشد، لازم هستند. همچنین فرض می‌کنیم $S(p)$ برای همه مقادیر $۰ < p \leq ۱$ تعریف شده و برای پیشامدهایی که $p = ۰$ است، تعریف نشده است.

اولین شرط، بیان این واقعیت است که وقتی بشنویم پیشامد حتمی که باید اتفاق می‌افتاده، رخ داده است، تعجب ما از وقوع آن صفر است، یعنی

اصل ۱. $S(۱) = ۰$.

دومین شرط، بیان می‌دارد که میزان تعجب از پیشامدی که شانس کمتری برای وقوع دارد، بیشتر از میزان تعجب برای پیشامدی با شانس وقوع بیشتر است.

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

---

۱) *Surprise*

سومین شرط، یک ویژگی ریاضی تابع $S(p)$ است، به‌طوری که انتظار می‌رود هر تغییر کوچک در p باعث تغییر کوچکی در $S(p)$ شود.

اصل ۳. $S(p)$ تابعی پیوسته از p است.

برای بیان آخرین شرط، دو پیشامد مستقل E و F را در نظر می‌گیریم، به طوری که $P(E)=p$ و $P(F)=q$ ، چون احتمال وقوع توأم برابر $P(EF)=pq$ است، پس میزان تعجب حاصل از وقوع توأم E و F برابر $S(pq)$ است. حال فرض کنیم، ابتدا مطلع شویم که E اتفاق افتاده است و پس از آن پیشامد F نیز اتفاق افتاده باشد، چون $S(p)$ میزان تعجب وقوع پیشامد E است، بنابراین $S(pq)-S(p)$ نشان دهنده افزایش تعجب است، وقتی مطلع شویم که F نیز اتفاق افتاده است. به علاوه چون E و F مستقل از هم هستند و اطلاع از وقوع یا عدم وقوع پیشامد E تاثیری در احتمال وقوع F ندارد، بنابراین افزایش تعجب بایستی دقیقاً برابر با $S(q)$ باشد.

اصل ۴. به ازای $۰<p\leq ۱$ و $۰<q\leq ۱$ ، $S(pq)=S(p)+S(q)$.

حال می‌توانیم قضیه زیر را که بیان کننده ساختار $S(p)$ است، مطرح کنیم.

۱-۲-۱ قضیه. تابع $S(p)$ در اصول ۱ تا ۴ 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. به مرجع [۳۴] رجوع شود.

۱-۲-۱ نکته. با توجه به این که C یک عدد حقیقی مثبت است، بنابراین برای سادگی محاسبات مقدار آن را معمولاً برابر یک قرار می‌دهند.

۱-۳ آنتروپی متغیرهای تصادفی گسسته

متغیر تصادفی X که مقادیر $x_1,\dots,x_n$ را با احتمالات متناظر $p_1,\dots,p_n$ انتخاب می‌کند، در نظر می‌گیریم. چون $-\log p_i$ نشان دهنده میزان تعجب حاصل از این است که متغیر تصادفی X مقدار x_i را اختیار کند، در نتیجه میانگین موزون (امید ریاضی) میزان تعجب حاصل از اطلاع در مورد مقدار متغیر تصادفی X برابر است با

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

کمیت $H(X)$ در نظریه اطلاع با عنوان آنتروپی متغیر تصادفی X شناخته می‌شود و برای اولین بار توسط شانون (۱۹۴۸) معرفی شد. در صورتی که یکی از مقادیر برابر صفر باشد، بازهم مشکلی پیش نمی‌آید، چرا که $\lim\limits_{u\to 0}u\log u=0$.

$H(X)$ نشان دهنده متوسط میزان تعجبی است که در مورد اطلاع از مقدار X حاصل می‌گردد، پس می‌توان آن را با عنوان نشان دهنده میزان عدم قطعیتی که برای مقدار X وجود دارد نیز تفسیر نمود. تعریف زیر به طور رسمی آنتروپی متغیر تصادفی گسسته را تعریف می‌کند.

۱-۳-۱ **تعریف.** فرض کنیم X یک متغیر تصادفی گسسته با تکیه‌گاه S_X و تابع احتمال $f_\theta(x)$ باشد، در این صورت آنتروپی X به صورت زیر تعریف می‌شود

$$\begin{aligned}H(X)&=-E\left(\log f_\theta(X)\right)\\&=-\sum_{x\in S_X}f_\theta(x)\log f_\theta(x)\end{aligned}$$

یادآوری این نکته ضروری است که آنتروپی X تابعی از احتمالات $f_\theta(x_i)=P_\theta(X=x_i)$ است و به مقادیر x بستگی ندارد، لذا به جای نماد $H(X)$ می‌توان از نماد $H(f(X))$ یا $H(f)$ و یا $H(p_1,p_2,\dots)$ استفاده کرد.

۱-۳-۱ **مثال.** آنتروپی متغیر تصادفی هندسی X با پارامتر p به صورت زیر به‌دست می‌آید

$$
\begin{aligned}
H(X) &= -\sum_{x=0}^{\infty} p(1-p)^x \log\left(p(1-p)^x\right) \\
&= -\sum_{x=0}^{\infty} p(1-p)^x \left(\log p + x\log(1-p)\right) \\
&= -\log p \sum_{x=0}^{\infty} p(1-p)^x - \log(1-p)\sum_{x=0}^{\infty} xp(1-p)^x \\
&= -\log p - \frac{1-p}{p}\cdot\log(1-p) = \frac{-p\log p-(1-p)\log(1-p)}{p} \\
&= \frac{\log\dfrac{(1-p)^{p-1}}{p^p}}{p} = \frac{h(p)}{p}
\end{aligned}
$$

که $h(p)$ آنتروپی توزیع برنولی با احتمال موفقیت p است.

می‌دانیم که اگر در $H(p_1, p_2, \ldots)$ احتمال‌ها را به ترتیب متفاوتی قرار دهیم، مقدار اطلاع تغییر نمی‌کند. مثال زیر نشان می‌دهد که توزیع‌های احتمال متفاوت می‌توانند به مقدار اطلاع یکسانی منجر گردند.

۱-۳-۲ مثال. دو توزیع احتمال زیر را در نظر می‌گیریم

$$
\begin{aligned}
P &= (0/5, 0/25, 0/25) \\
Q &= (0/48, 0/32, 0/20)
\end{aligned}
$$

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

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

پس $H(p)$ تابع توزیع را به طور یکتا مشخص نمی‌کند، یعنی توزیع‌های احتمال متفاوت می‌توانند به مقدار اطلاع یکسانی منجر گردند. برخی از آزمایش‌ها می‌توانند توزیع‌های احتمالی متفاوتی داشته باشند ولی مقدار اطلاع آنها یکسان باشد. در فصل‌های دوم و سوم روی این مطلب برای توزیع‌های طول عمر بحث خواهیم کرد و شرایط لازم را برای این که آنتروپی گذشته و آنتروپی گذشته تعمیم یافته تابع توزیع طول عمر را مشخص کنند، تعیین می‌کنیم.

به صورت شهودی $H(f)$ عبارت است از عدم قطعیت موجود در f برای قابل پیش بینی بودن برآمدی از X. آنتروپی یکنواختی[1] یک توزیع را اندازه می‌گیرد. با افزایش $H(f)$ ، $f(x)$ به توزیع یکنواخت نزدیک‌تر می‌شود، در نتیجه تراکم احتمال‌ها کمتر شده و پیش‌بینی برآمدی از X مشکل‌تر می‌شود. درواقع آنتروپی توزیع‌های نوک‌تیز[2] پایین است، در حالی که اگر احتمال پخش شده باشد آنتروپی خیلی بالاتر است. به عبارت دیگر $H(f)$ متراکم بودن احتمال‌ها را اندازه می‌گیرد. توزیع‌های با آنتروپی پایین بسیار متراکم‌تر هستند و بنابراین قدرت پیش‌بینی ما درمورد این گونه توزیع‌ها بیشتر است.

در تعریف ۱-۳-۱ پایه لگاریتم تعیین نشده است. در اکثر موارد مهم نیست که آن را چند قرار دهیم، چرا که تغییر در پایه لگاریتم صرفاً تغییر در مقیاس واحدها می‌باشد، اما اکثراً پایه را ۲ یا e (عدد نپر) قرار می‌دهند. واحد اطلاع را هنگامی که در پایه ۲ محاسبه شود، bit (دستگاه دومقداری) می‌نامند و هنگامی که در پایه e محاسبه شود، واحد اطلاع برحسب nat[3] (واحد طبیعی) بیان می‌شود. به سادگی می‌توان دید که رابطه بین bit و nat به صورت زیر است

$$۱\,bit = ۰/۶۹۳\,nat$$

در طول فصل اول از نماد log برای نشان دادن لگاریتم در مبنای دو و از نماد ln برای نشان دادن لگاریتم طبیعی استفاده می‌کنیم، در حالی که در بقیه فصول برای سادگی محاسبات فقط از واحد طبیعی استفاده می‌کنیم و از نماد log برای نشان دادن لگاریتم طبیعی استفاده کرده‌ایم.

۱-۳-۳ مثال. فرض کنیم X متغیر تصادفی برنولی با احتمال موفقیت p باشد، آنگاه نمودار آنتروپی این متغیر تصادفی بر حسب تنها پارامترش، p ، به صورت زیر می‌باشد

---
۱) *Uniformly*

۲) *Sharply peaked distributions*

۳) *Natural*