دانشگاه تربیت معلم

دانشکده علوم ریاضی و کامپیوتر

پایان‌نامه کارشناسی ارشد

آمار (گرایش آمار ریاضی)

عنوان :

نتایجی راجع به آنتروپی باقی‌مانده تعمیم‌یافته

تدوین :

علی علی‌پور

استاد راهنما :

دکتر علی‌اکبر رحیم‌زاده ثانی

خرداد ۱۳۸۹

Tarbiat Moallem University

Faculty of Mathematical Sciences and Computer

Statistics (Mathematics Statistics)

Thesis Submitted of the Degree of

Master of Science in Statistics

Title :

Some Results on Generalized Residual Entropy

By :

Ali Alipour

Supervisor :

Dr. A. A. Rahimzadeh sani

June ٢٠١٠

چکیده :

آنتروپی شانون نقش مهمی در زمینه نظریه اطلاع دارد. چون آنتروپی شانون در سیستم‌هایی که برای مدتی معین فعال بوده‌اند کاربرد ندارد، مفهوم آنتروپی باقی‌مانده در تحقیقات توسعه داده شده است. ما آنتروپی باقی‌مانده را با انتخاب یک تابع محدب ϕ به طوری‌که $\phi(۱) = ۰$ تعمیم می‌دهیم. در این پایان نامه، بعضی از ویژگی های ترتیب و طول عمر برحسب تابع آنتروپی باقی‌مانده تعمیم‌یافته تعریف شده و خاصیت‌های آن‌ها مطالعه شده است. تقریباً جزئیات موجود کمی در تحقیقات تعمیم داده شده است و بعضی توزیع‌ها (از جمله : یکنواخت، نمایی، پارتو و برد متناهی) به واسطه آنتروپی باقی‌مانده تعمیم‌یافته مشخص شده‌اند.

واژه‌های کلیدی : فاصله اطلاع جهت‌دار، کلاس $DURL$ و کلاس $IURL$، اندازه اطلاع، تعمیر اولیه، آنتروپی باقی‌مانده، سیستم بهبود و تباهی.

رده‌بندی موضوعی : آنتروپی؛ ۵۴C۹۹,۵۴C۷۰.

فهرست مطالب

مقدمه

شانون[1] (۱۹۴۸) اولین کسی بود که مفهوم آنتروپی را مطرح کرد، که در نظریه اطلاع به آنتروپی شانون یا نظریه اطلاع شانون معروف است[22]. خینچن[2] (۱۹۵۷) آنتروپی شانون را با انتخاب یک تابع محدب ϕ، به طوری که $\phi(1)=0$ تعمیم داد[17].ناندا و پائول[3] (۲۰۰۶) توابع آنتروپی تعمیم‌یافته را تعریف نمودند[21]. آنتروپی شانون نقشی مهم در زمینه نظریه اطلاع دارد، اما چون آنتروپی شانون در سیستمهایی که برای مدتی معین فعال بوده‌اند کاربرد ندارد، مفهوم آنتروپی باقی‌مانده در تحقیقات توسعه داده شده است. ابراهیمی[4] (۱۹۹۶) آنتروپی باقی‌مانده را برای سیستمهایی که برای مدتی معین فعال بوده‌اند، تعریف کرد[11]. ناندا و پائول (۲۰۰۶) آنتروپی باقی‌مانده تعمیم‌یافته را تعریف نمودند[21]. ابراهیمی (۱۹۹۶) دو کلاس از طول عمر توزیع‌ها (DURL و IURL) را تعریف کرد[11]. ناندا و پائول (۲۰۰۶) دو کلاس جدید از طول عمر توزیع‌ها براساس توابع آنتروپی باقی‌مانده تعمیم‌یافته را تعریف کردند[21]. ابراهیمی و کرمانی[5] (۱۹۹۶) تعریف اندازه کالبک-لایبلر[6] را به منظور تطبیق نمودن با عمر فعلی سیستم، اصلاح کردند[13]. گاپتا و ناندا[7] (۲۰۰۲) نوع جدیدی از اندازه اطلاع را به ازای دو انتخاب متمایز ϕ در رابطه (۹.۱) تعریف کردند، به طوری که عمر فعلی سیستم قابل محاسبه است[16]. به طور کلی مطالبی که در این پایان‌نامه ارائه می‌شود، به شرح زیر است :

فصل اول : تعاریف و مقدماتی از آنتروپی

در این فصل ابتدا در بخش اول، مقدمه‌ای از آنتروپی خواهیم گفت. در بخش دوم، میزان تعجب حاصل از وقوع یک پیشامد را کمی می‌کنیم و با استفاده از یک قضیه‌ی مهم، اندازه‌ای برای عدم قطعیت متغیرهای تصادفی گسسته به دست می‌آوریم. در بخش سوم نیز آنتروپی متغیرهای تصادفی گسسته و به دنبال آن تعریف آنتروپی توام و شرطی را ارائه خواهیم کرد، که درآن با استفاده از چند قضیه، برخی از خواص آنتروپی مورد بحث و بررسی قرار گرفته است. در بخش آخر نیز آنتروپی متغیرهای تصادفی

۱) Shannon,
۲) Khinchin,
۳) Nanda & *Paul*,
٤) Ebrahimi,
۵) Ebrahimi & *Kirmani*,
٦) Kullback – Leibler,
۷) Gupta & *Nanda*.

پیوسته و تعریف آنتروپی شانون برای متغیرهای تصادفی نامنفی مطلقاً پیوسته را بیان می‌کنیم و به دنبال آن تعاریفی از آنتروپی تعمیم‌یافته، آنتروپی باقی‌مانده و نیز آنتروپی باقی‌مانده تعمیم‌یافته خواهیم آورد.

فصل دوم : رابطه‌های ترتیبی براساس آنتروپی باقی‌مانده تعمیم‌یافته

در این فصل در بخش اول مقدمه‌ای بیان خواهیم کرد. در بخش دوم تعاریفی از رابطه‌های ترتیبی بیان کرده و سپس یک لم بسیار مفید بیان خواهیم کرد. در بخش سوم آنتروپی باقی‌مانده تعمیم‌یافته تحت تبدیل خطی صعودی خواهد آمد، که قضایا و نتایج سودمندی براساس آن بیان خواهیم کرد. در بخش آخر نیز ترتیب براساس آنتروپی باقی‌مانده تعمیم‌یافته در متغیر تصادفی نمایی دوپارامتری ارائه خواهد شد و سپس فرآیند تصادفی تولید شده با روش تعمیرهای اولیه خواهد آمد.

فصل سوم : کلاسهای جدید طول عمر بر اساس آنتروپی باقی‌مانده تعمیم‌یافته

در این فصل در بخش اول ابتدا تعریفی از کلاسهای طول عمر باقی‌مانده با عدم قطعیت صعودی (نزولی) و سپس تعاریفی براساس نرخ خرابی و میانگین طول عمر باقی‌مانده بیان خواهیم کرد. در بخش دوم راجع به یکنوایی توزیع‌ها بر اساس آنتروپی باقی‌مانده تعمیم‌یافته بحث خواهیم کرد، مثال نقضی ارائه خواهیم کرد که نشان می‌دهد همه توزیع‌ها بر اساس آنتروپی باقی‌مانده تعمیم‌یافته یکنوا نیستند. براساس قضیه‌ای نشان خواهیم داد که توزیع نمایی تنها توزیعی است که برحسب عدم قطعیت باقی‌مانده در طول عمر نوع اول و دوم، صعودی و نیز نزولی است. در بخش آخر ابتدا قضیه‌ای بیان خواهیم کرد که نشان می‌دهد کلاسهای مذکور بر اساس آنتروپی باقی‌مانده تعمیم‌یافته تحت تبدیل خطی صعودی بسته هستند. سپس مثالی از کاربرد این قضیه خواهیم آورد و بالاخره براساس قضیه‌ای کران‌های بالا و پایین برای توابع آنتروپی باقی‌مانده تعمیم‌یافته ارائه می‌دهیم.

فصل چهارم : مشخصه‌سازی برخی از توزیع‌های پیوسته و گسسته برحسب آنتروپی باقی‌مانده تعمیم‌یافته

در این فصل ابتدا در بخش اول، مقدمه‌ای بیان خواهیم کرد. در بخش دوم، ابتدا قضیه‌ای بیان خواهیم کرد که نشان می‌دهد در صورت نزولی بودن توابع آنتروپی باقی‌مانده تعمیم‌یافته، تابع توزیع به طور یکتا مشخص می‌شود. به دنبال آن توزیع یکنواخت پیوسته و توزیع نمایی را برحسب توابع آنتروپی باقی‌مانده

تعمیم‌یافته مشخصه‌سازی می‌کنیم. در بخش سوم نیز، با بیان دو قضیه توزیع‌های نمایی، پارتو و برد متناهی را برحسب توابع آنتروپی باقی‌مانده تعمیم‌یافته مشخصه‌سازی می‌کنیم. در بخش چهارم، به مشخصه‌سازی مدل نرخ مخاطره نسبی (PHR) کاکس برحسب توابع آنتروپی باقی‌مانده تعمیم‌یافته می‌پردازیم. در بخش آخر نیز نتایجی از توزیع‌های گسسته برحسب آنتروپی مذکوربیان می‌کنیم و سرانجام توزیع گسسته یکنواخت را برحسب آنتروپی مذکور مشخصه‌سازی می‌کنیم. این پایان‌نامه براساس مقاله زیر تدوین شده است :

Asok K. Nanda, Prasanta Paul, Some results on generalized residual entropy, Information Science 176 (2006) 27-47.

فصل ۱

تعاریف و مقدماتی از آنتروپی

۱.۱ مقدمه

امروزه آنتروپی و نظریه‌ی اطلاع نقشی اساسی در رشته‌های مختلفی همچون آمار، مهندسی مکانیک، برق، نجوم، اقتصاد، پزشکی، مدیریت و سایر رشته‌ها دارد.

آنتروپی از واژه یونانی *Entropy* به معنی ((به درون خود می‌روم)) گرفته شده است. این اصطلاح برای اولین بار توسط کلاوسیوس[۱] فیزیکدان آلمانی (۱۸٦٥) در زمینه ترمودینامیک به کار گرفته شد[٤].

در واقع می‌توان گفت که آنتروپی شانون (۱۹٤۸) هسته‌ی اصلی نظریه‌ی اطلاع را تشکیل می‌دهد که گاهی اوقات تحت عنوان اندازه‌ی عدم قطعیت به آن مراجعه می‌کنیم[22]. آنتروپی به ندرت در کتابهای آماری مورد بحث و بررسی قرار گرفته است. آنتروپی موضوعی است که به نوعی به عدم قطعیت و اطلاع مربوط بوده و در رابطه با ترمودینامیک آماری و رمزگذاری به‌کار می‌رود. آنتروپی را می‌توان با بی نظمی معادل دانست، یعنی هر چه نظم سیستمی بالا رود آنتروپی آن کاهش می‌یابد و بالعکس کاهش نظم باعث افزایش آنتروپی می‌شود.

آنتروپی یک سیستم با میزان اطلاع موجود در آن مرتبط است، یعنی یک سیستم با نظم بیشتر می‌تواند با بایت‌های کمتری از اطلاعات توصیف شود در حالی که سیستمی با بی‌نظمی بالاتر برای توصیف شدن به

۱) *Clausius*

بایت‌های بیشتری از اطلاعات نیازمند است. نظریه‌ی اطلاع در ابتدا برای بیان عددی اطلاع به وجود آمد. همان طورکه فاصله، دما، زمان، و ... را با عدد اندازه‌گیری می‌کنند، مقدار اطلاعی را که یک موضوع به ما می‌دهد، به وسیله تعداد سوالات مثبت لازم برای پی بردن به آن موضوع اندازه‌گیری می‌شود. این جواب-های به دست 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) می‌نامیم. اگر موضوعی که مورد نظر است در فضای غیر هم-شانس قرار داشته باشد متوسط تعداد سوال‌هایی را که برای رسیدن به موضوع لازم است، اطلاع شانون (آنتروپی شانون) گوئیم و با $H(X)$ یا $H(f)$ نشان می‌دهیم. به منظور درک بهتری از آنتروپی به عنوان یک اندازه برای کاهش بایت‌های لازم برای مخابره یک پیام مثال زیر را در نظر می‌گیریم.

مثال ۱.۱.۱ فرض کنیم مقدار بردار تصادفی گسسته X در موقعیت فرستنده A مشاهده شده و توسط یک شبکه مخابراتی که شامل دو سیگنال صفر ویک 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 هر گونه ابهامی، لازم است که هیچ دنباله کدگذاری شده را نتوان با استفاده از دنباله کوتاه‌تری به وسیله اضافه نمودن اجزای بیشتر به دست آورد. به عنوان مثال اگر X بتواند مقادیر $x_۱$، $x_۲$، $x_۳$ و $x_٤$ را اختیار کند، یکی از روش‌های کدگذاری عبارتست از

$$x_۱ \leftrightarrow ۰۰,$$

$$x_۲ \leftrightarrow ۰۱,$$

$$x_۳ \leftrightarrow ۱۰,$$

$$x_٤ \leftrightarrow ۱۱.$$

اگر $X = x_۱$، آنگاه پیام ۰۰ به موقعیت گیرنده B ارسال می‌شود و اگر $X = x_۲$ باشد، پیام ۰۱ به موقعیت گیرنده B ارسال می‌شود و ... .

حال احتمال‌های زیر را در نظر می‌گیریم

$$P\{X = x_۱\} = \frac{۱}{۲},$$

$$P\{X = x_۲\} = \frac{۱}{٤},$$

$$P\{X = x_۳\} = P\{X = x_٤\} = \frac{۱}{۸}.$$

لذا برای کد فوق انتظار داریم به طور متوسط به ازای هر مقدار ممکن X، دو بیت ارسال شود :

$$[(۰/۵ \times ۲ + ۰/۲۵ \times ۲ + ۰/۱۲۵ \times ۲ + ۰/۱۲۵ \times ۲) = ۲].$$

با استفاده از ویژگی مفید نظریه‌ی اطلاع می‌توان یک کدگذاری بهتر یافت که به بایت‌های کمتری نیاز داشته باشد. به عنوان مثال، کدگذاری به صورت زیر را در نظر می‌گیریم

$$\left.\begin{array}{ll} x_۱ \leftrightarrow ۰ & ۰/۵ \times ۱ = ۰/۵ \\ x_۲ \leftrightarrow ۱۰ & ۰/۲۵ \times ۲ = ۰/۵ \\ x_۳ \leftrightarrow ۱۱۰ & ۰/۱۲۵ \times ۳ = ۰/۳۷۵ \\ x_٤ \leftrightarrow ۱۱۱ & ۰/۱۲۵ \times ۳ = ۰/۳۷۵ \end{array}\right\}$$

یعنی به طور متوسط ۱/۷۵ بیت به ازای هر مقدار X.

علاوه بر این، نظریه اطلاع به ما می‌گوید که آنتروپی این منبع اطلاعات ۱/۷۵ بیت به ازای هر مقدار X می‌باشد. لذا این کدگذاری نسبت به سایر کدگذاری‌ها برای توصیف شدن این منبع اطلاعاتی بهترخواهد بود.

۲.۱ تعجب (عدم قطعیت یا آنتروپی)

پیشامد A را در نظر می گیریم که می‌تواند رخ دهد، یعنی احتمال وقوع آن صفر نباشد، اگر آزمایش انجام گیرد و گفته شود که A رخ داده است، چه قدر تعجب می‌کنیم؟ به نظر منطقی به نظر می‌رسد که تصور کنیم میزان تعجب حاصل از وقوع A تنها به احتمال وقوع A بستگی داشته باشد.

برای مثال فرض کنیم آزمایش شامل انتخاب دو عدد به طور تصادفی و با جایگذاری از مجموعه $\{۱,۲,۳,٤,۵\}$ باشد، اگر بشنویم که مجموع اعداد ظاهر شده زوج است زیاد متعجب نمی‌شویم چون احتمال وقوع آن برابر $\frac{۱۳}{۲۵}$است، در حالی‌که اگر بشنویم مجموع اعداد ظاهرشده ۲ است بیشتر تعجب می کنیم، چرا که احتمال وقوع آن برابر $\frac{۱}{۲۵}$ 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 A تنها بستگی به احتمال وقوع آن دارد و آن را با $h(p)$ نمایش می‌دهیم، یعنی

میزان تعجبی که از وقوع پیشامدی با احتمال p حاصل می‌گردد. حال سعی می‌کنیم که شکل تابعی $h(p)$ را بر اساس توافق روی مجموعه‌ای از شرایط منطقی تعیین نموده و سپس ثابت کنیم که این اصول برای این که $h(p)$ شکل معینی داشته باشد، لازم هستند. همچنین فرض می‌کنیم $h(p)$ برای همه مقادیر $0 < p \leq 1$ تعریف شده و برای پیشامدهایی‌که $p = 0$، تعریف نشده است.

شرط اول بیانگر 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. $h(1) = 0$.

شرط دوم، بیان می‌کند که میزان تعجب از پیشامدی که شانس کمتری برای وقوع دارد، بیشتر از میزان تعجب برای پیشامدی با شانس وقوع بیشتر است، یعنی

اصل۲. $h(p)$ تابعی اکیداً نزولی از p است، یعنی اگر $p < q$، آن‌گاه $h(p) > h(q)$.

شرط سوم یک ویژگی ریاضی تابع $h(p)$ است، به طوری که انتظار می‌رود هر تغییر کوچک در p باعث تغییر کوچکی در $h(p)$ شود، یعنی

اصل ۳. $h(p)$ تابعی پیوسته از p است.

شرط چهارم، دو پیشامد مستقل A و B را در نظر می‌گیریم به طوری‌که $P(A) = p$ و $P(B) = q$، چون احتمال وقوع توام A و B به دلیل استقلال برابر $P(AB) = pq$ است، پس میزان تعجب حاصل از وقوع توام A و B برابر $h(pq)$ است. حال فرض کنیم، ابتدا مطلع شویم که A رخ داده و پس از آن پیشامد B نیز رخ داده است، چون $h(p)$ میزان تعجب وقوع پیشامد A است، لذا $h(pq) - h(p)$ نشان دهنده افزایش تعجب است، وقتی که مطلع شویم B نیز رخ داده است. به علاوه چون A و B از هم مستقل هستند و اطلاع از وقوع یا عدم وقوع پیشامد A تاثیری در احتمال وقوع پیشامد B ندارد، افزایش تعجب بایستی دقیقاً برابر با $h(q)$ باشد، یعنی $h(pq) - h(p) = h(q)$. لذا

اصل٤. به ازای $0 < p \leq 1$ و $0 < q \leq 1$،

$$h(pq) = h(p) + h(q).$$

اصول فوق عدم قطعیت را به طور کامل مشخص می‌کنند.

حال قضیه‌ی زیر را که بیان کننده ضابطه تابع $h(p)$ است، مطرح می‌کنیم.

قضیه ۱.۲.۱ تابع $h(p)$ در اصول فوق صدق می‌کند، اگر و فقط اگر $h(p) = -cE(\log p)$ که c یک عدد دلخواه صحیح و مثبت ومبنای لگاریتم همواره بزرگتر از یک است.

برهان : برهان در مرجع [۱] آمده است.

تذکر ۱.۲.۱ چون c یک عدد حقیقی مثبت است، لذا برای سادگی محاسبات مقدار آن را معمولاً برابر یک قرار می‌دهند.

۳.۱ آنتروپی متغیرهای تصادفی گسسته

تعریف ۱.۳.۱ اگر X یک متغیر تصادفی گسسته با مقادیر $x_۱$، $x_۲$، ... ، x_n، ... و با احتمالهای به ترتیب $p_۱$، $p_۲$، ... ، p_n، ... باشد، در این صورت آنتروپی X به صورت زیر تعریف می‌شود :

$$H(p) = H(p_۱\,, p_۲\,, \ldots, p_n\,,\ \ldots) = -\sum_i p_i \ln p_i$$

به شرط آن که سری فوق همگرا باشد.

تذکر ۱.۳.۱ اگر یکی از p_iها برابر صفر باشد، طبق قرداد $p_i \log p_i$ را برابر صفر در نظر می‌گیریم.

تذکر ۲.۳.۱ 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 $H\big(f(X)\big)$، $H(f)$ و یا $H(p_۱, p_۲, \ldots)$ استفاده کرد.

تذکر ۳.۳.۱ اگر X در تعریف ۱.۳.۱، مقادیر متناهی $x_۱$ ، $x_۲$ ، ... و x_n را با احتمالهای به ترتیب $p_۱$ ، $p_۲$ ، ... ، p_n اختیار کند، در این صورت آنتروپی X به صورت زیر تعریف می‌شود :

$$H(p) = H(p_۱, p_۲, \ldots, p_n) = -\sum_{i=۱}^{n} p_i \ln p_i\,. \qquad (۱.۳.۱)$$

اگر i ای باشد که $p_i = ۱$، یعنی، متغیر تصادفی X مقدار x_i را با اطمینان بگیرد، آنگاه $H(p) = ۰$. این یعنی وقتی که $H(p) = ۰$، عدم قطعیت درباره پیش‌بینی‌پذیری X به وسیله تابع جرم p وجود ندارد. به طور مشابه وقتی که X دارای توزیع یکنواخت گسسته است، یعنی، برای هر i، $p_i = \frac{۱}{n}$، این یک حالت نادر است زیرا پیش‌بینی نتیجه یک چنین آزمایشی سخت است. یک ایراد (۱.۳.۱) این است که حتی اگر p_i و p_j متمایز باشند، ممکن است اطلاع حاصل از x_i و x_j یکسان باشند(به مرجع [5] مراجعه شود). برای نشان دادن این موضوع، فرض کنیم $n = ۳$، $p_۱ = p_۲ = \frac{۱}{۴}$ و $p_۳ = \frac{۱}{۲}$. در این صورت سهم $x_۱$، $\frac{۱}{۴} ln ۴$ و سهم $x_۳$، $\frac{۱}{۲} ln ۲$ است که با هم برابرند اگر چه $p_۱ \neq p_۳$. برای رفع این اشکال آنتروپی‌های گسسته تعمیم‌یافته در فصل ٤ تعریف شده و خاصیتهای آنها مورد مطالعه قرار گرفته است.

مثال ۱.۳.۱ فرض کنیم

$$P(X = ۱) = p \quad , \ P(X = ۰) = ۱ - p$$

در این‌صورت

$$H(X) = -p \log p - (۱ - p) \log(۱ - p) =: H(p).$$

اگر $p = \frac{۱}{۲}$ آنگاه، بیت $H(X) = ۱$، اگر ۱ یا $p = ۰$، $H(X)$ برابر صفر است. این واضح است، زیرا هنگامی‌که ۱ یا $p = ۰$، عدم قطعیت وجود ندارد. به طور مشابه هنگامی که $p = \frac{۱}{۲}$ عدم قطعیت ماکسیمم است، که همچنین برابر ماکسیمم مقدار آنتروپی است.

تذکر ٤.۳.۱ در اکثر موارد مهم نیست که پایه لگاریتم 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 می‌دهیم. واحد اطلاع را هنگامی‌که در پایه ۲ محاسبه شود، بیت (مخفف اعداد دودویی) و هنگامی که در پایه e محاسبه شود، نات (واحد طبیعی) می‌نامند.

رابطه‌ی بین bit و nat به صورت زیر است :

$$۱ \ bit = ۰/٦۹۳ \ nat .$$

در فصل اول از نماد log برای نشان دادن لگاریتم در مبنای ۲ و از نماد ln برای نشان دادن لگاریتم طبیعی استفاده می‌کنیم. اما در بقیه فصول فقط از نماد ln استفاده می‌کنیم.

مثال ۲.۳.۱ آنتروپی متغیر تصادفی هندسی X با پارامتر p $(f(x) = p(۱-p)^x)$ به صورت زیر به دست می‌آید :

$$
\begin{aligned}
H(X) &= -\sum_{x=۰}^{\infty} p(۱-p)^x \, log(p(۱-p)^x) \\
&= -\sum_{x=۰}^{\infty} p(۱-p)^x \, (log\, p + x\, log(۱-p)) \\
&= -log\, p \, \sum_{x=۰}^{\infty} p(۱-p)^x - log(۱-p) \, \sum_{x=۰}^{\infty} x\, p(۱-p)^x \\
&= -log\, p - \frac{۱-p}{p} log(۱-p) \\
&= \frac{-p\, log\, p - (۱-p)\, log(۱-p)}{p} = \frac{h(p)}{p}.
\end{aligned}
$$

که $h(p)$ آنتروپی توزیع برنولی است.

لم ۱.۳.۱ برای هر $x > ۰$، داریم : $ln\, x \leq x - ۱$ و تساوی زمانی برقرار است که $x = ۱$.

برهان : برهان در مرجع [۳] آمده است.

تعریف ۲.۳.۱ اگر X یک متغیر تصادفی با تابع چگالی یا احتمال f باشد، آن‌گاه تکیه‌گاه X عبارتست از :

$$\{x \in \mathbb{R} \mid f(x) > ۰\}.$$

قضیه ۱.۳.۱ فرض کنیم $S_x = \{x_۱, \ldots, x_n\}$ تکیه‌گاه متغیر تصادفی X باشد و مولفه‌های بردار احتمال $P = \{p_۱, \ldots, p_n\}$ به ترتیب احتمالهای وابسته به پیشامدهای $\{X = x_i\}$ باشند، در این صورت

الف) $H(X) \geq ۰$؛

ب) آنتروپی X صفر است اگر و تنها اگر X در وضعیت قطعیت کامل باشد (مقداری ثابت را با احتمال یک بگیرد)؛

ج) آنتروپی از بالا کراندار بوده و حداکثر آنتروپی هنگامی است که X دارای توزیع یکنواخت باشد. به عبارت دیگر $H(X) \leq log\, n$ و تساوی برقرار است اگر و تنها اگر $p_۱ = \cdots = p_n = \frac{۱}{n}$.

برهان : برهان در مرجع [10] آمده است.

تعریف ۳.۳.۱ (آنتروپی متغیرهای تصادفی توام) آنتروپی توام دو متغیر تصادفی (X,Y) با توزیع توام (تابع احتمال توام) $P(X,Y)$ به صورت زیر تعریف می‌شود :

$$H(X,Y) = -E(log\,P(X,Y)) = -\sum_i \sum_j P\left(x_i, x_j\right) log\,P\left(x_i, x_j\right)$$

به شرط آن که سری فوق همگرا باشد.

تعریف ٤.۳.۱ (آنتروپی شرطی) اگر $(X,Y) \sim P(X,Y)$، آنگاه آنتروپی شرطی $H(Y|X)$ برابر است با

$$\begin{aligned} H(Y|X) &= -\sum_i P(x_i)\, H(Y|X = x_i) \\ &= -\sum_i P(x_i) \sum_j P\left(y_j|x_i\right) log\,P\left(y_j|x_i\right) \\ &= -\sum_i \sum_j P\left(x_i, y_j\right) log\,P\left(y_j|x_i\right) \\ &= -E(log\,P(Y|X)) \end{aligned}$$

برای بیش از دو متغیر تصادفی داریم :

$$H(X,Y|Z) = -\sum_i \sum_j \sum_k P\left(X_i, Y_j, Z_k\right) log\,P\left(X_i, Y_j|Z_k\right),$$
$$H(X|Y,Z) = -\sum_i \sum_j \sum_k P\left(X_i, Y_j, Z_k\right) log\,P\left(X_i|Y_j, Z_k\right),$$

به شرط آن که سریهای فوق همگرا باشند.

قضیه ۲.۳.۱ (قانون زنجیری) برای دو متغیر تصادفی X و Y همواره داریم :

$$H(X,Y) = H(X) + H(Y|X).$$

برهان : برهان در مرجع [10] آمده است.

نتیجه ۱.۳.۱ برای متغیرهای تصادفی X، Y و Z همواره داریم :

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z)\,.$$

برهان : برهان در مرجع [10] آمده است.

تذکر ٤.۳.۱ توجه کنیم که درحالت کلی $H(Y|X) \neq H(X|Y)$، اگر چه

$$H(X)-H(X|Y)=H(Y)-H(Y|X).$$

مثال زیر بیانگر این واقعیت است.

مثال ۳.۳.۱ فرض کنیم (X,Y) دارای توزیع توام زیر باشند :

$Y \backslash X$	۱	۲	۳	٤
۱	$\frac{۱}{۸}$	$\frac{۱}{۱٦}$	$\frac{۱}{۳۲}$	$\frac{۱}{۳۲}$
۲	$\frac{۱}{۱٦}$	$\frac{۱}{۸}$	$\frac{۱}{۳۲}$	$\frac{۱}{۳۲}$
۳	$\frac{۱}{۱٦}$	$\frac{۱}{۱٦}$	$\frac{۱}{۱٦}$	$\frac{۱}{۱٦}$
٤	$\frac{۱}{٤}$	۰	۰	۰

جدول (۱.۳.۱) توزیع توام X و Y مربوط به مثال ۳.۳.۱

توزیع حاشیه‌ای X، $\left(\frac{۱}{۲},\frac{۱}{٤},\frac{۱}{۸},\frac{۱}{۸}\right)$ و توزیع حاشیه‌ای Y، $\left(\frac{۱}{٤},\frac{۱}{٤},\frac{۱}{٤},\frac{۱}{٤}\right)$ است، لذا $H(X)=\frac{۷}{٤}$ و $H(Y)=۲$. همچنین $H(X|Y)=\frac{۱۱}{۸}$، $H(Y|X)=\frac{۱۳}{۸}$ و $H(X,Y)=\frac{۲۷}{۸}$ و می‌بینیم که $H(X|Y)\neq H(Y|X)$.

قضیه ۳.۳.۱ برای دو متغیر تصادفی X و Y همواره داریم :

الف) $H(X|Y)\leq H(X)$ و تساوی زمانی برقرار است که Xو Y مستقل باشند؛

ب) $H(X,Y)\leq H(X)+H(Y)$ و تساوی زمانی برقرار است که X و Y مستقل باشند.

برهان : برهان در مرجع [10] آمده است.

نتیجه ۲.۳.۱ برای هر n متغیر تصادفی $X_۱$، $X_۲$، ... و X_n داریم :

$$H(X_۱,X_۲,\dots,X_n)\leq \sum_{i=۱}^{n} H(X_i).$$

برهان : برهان در مرجع [10] آمده است.

کاربرد تعاریف و قضایای فوق را در مثال زیر مشاهده می‌کنیم.

مثال ۴.۳.۱ خانواده‌ای دارای ۲ فرزند است. از جنسیت این دو فرزند اطلاعی در دست نیست. مقایسه پاسخ سوالات زیر جالب خواهد بود.

۱) اندازه‌ی عدم قطعیت جنسیت فرزندان این خانواده .

۲) اندازه‌ی عدم قطعیت جنسیت فرزند اول و دوم .

۳) اگر بدانیم فرزند بزرگتر پسر است، اندازه‌ی عدم قطعیت چه قدر کاهش می‌یابد؟

۴) اگر بدانیم یکی از فرزندان پسر است، اندازه‌ی عدم قطعیت چه قدر کاهش می‌یابد؟

حل : فرض کنیم X و Y به ترتیب نشان دهنده جنسیت فرزند اول و دوم باشند. اگر پسر را با یک و دختر را با صفر نشان دهیم، در این صورت $S_{XY}=\{(0,0),(0,1),(1,0),(1,1)\}$. با فرض هم‌شانس بودن فرزند پسر و دختر در این خانواده، جدول توزیع توام احتمال X و Y به صورت زیر است :

$Y \backslash X$	0	1	$f_Y(y)$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$f_X(x)$	$\frac{1}{2}$	$\frac{1}{2}$	1

جدول (۲.۳.۱) توزیع توام X و Y مربوط به مثال ۴.۳.۱

بنابراین،

۱) $$H(X,Y)=-\left(\frac{1}{4}log\frac{1}{4}+\frac{1}{4}log\frac{1}{4}+\frac{1}{4}log\frac{1}{4}+\frac{1}{4}log\frac{1}{4}\right)=-\,log\frac{1}{4}=2$$

۲) $$H(X)=-\left(\frac{1}{2}log\frac{1}{2}+\frac{1}{2}log\frac{1}{2}\right)=-log\frac{1}{2}=1$$

$$H(Y) = -\left(\tfrac{۱}{۲} log \tfrac{۱}{۲} + \tfrac{۱}{۲} log \tfrac{۱}{۲}\right) = -\, log \tfrac{۱}{۲} = ۱$$

۳) $$H(Y|X = ۱) = -\left(\tfrac{۱}{۲} log \tfrac{۱}{۲} + \tfrac{۱}{۲} log \tfrac{۱}{۲}\right) = -log \tfrac{۱}{۲} = ۱$$

$$H(X,Y) - H(Y|X = ۱) = (۲ - ۱) = ۱$$

لذا با دانستن این‌که فرزند اول پسر است، به اندازه ۱ بیت از عدم قطعیت کاسته می‌شود.

٤) اگر پیشامد پسر بودن یکی از فرزندان را با B نشان دهیم، در اینصورت $B = \{(۰,۱), (۱,۰), (۱,۱)\}$ و در این حالت جدول توزیع توام احتمال به صورت زیر است :

Y \ X	۰	۱	$f(y\|B)$
۰	۰	$\frac{۱}{۳}$	$\frac{۱}{۳}$
۱	$\frac{۱}{۳}$	$\frac{۱}{۳}$	$\frac{۲}{۳}$
$f(x\|B)$	$\frac{۱}{۳}$	$\frac{۲}{۳}$	۱

جدول (۳.۳.۱) توزیع توام X و Y مربوط به مثال ٤.۳.۱

لذا،

$$H(X,Y|B) = -\left(۰ + \tfrac{۱}{۳} log \tfrac{۱}{۳} + \tfrac{۱}{۳} log \tfrac{۱}{۳} + \tfrac{۱}{۳} log \tfrac{۱}{۳}\right) = -\, log \tfrac{۱}{۳} = ۱/۵۸۵;$$

$$H(X,Y) - H(X,Y|B) = (۲ - ۱/۵۸۵) = ۰/٤۱۵.$$

بنابراین با دانستن این‌که یکی از فرزندان پسر است به اندازه‌ی ۰/٤۱۵ بیت از عدم قطعیت کاسته می‌شود.

اکنون از مثال فوق نتایج زیر حاصل می‌شود :

الف) به دلیل استقلال X و Y، $H(X,Y) = H(X) + H(Y)$.

ب) با مقایسه جوابهای (۳) و (٤) مشاهده می‌شود که اندازه کاهش در عدم قطعیت جنسیت فرزندان با اطلاع کسب شده توسط بند (۳)، ۱ بیت و با اطلاع کسب شده توسط بند (٤)، ۰/٤۱۵ بیت می‌باشد. این امر دور از انتظار نیست، زیرا بند (۳) اطلاع بیشتری از جنسیت فرزندان این خانواده به دست می‌دهد.