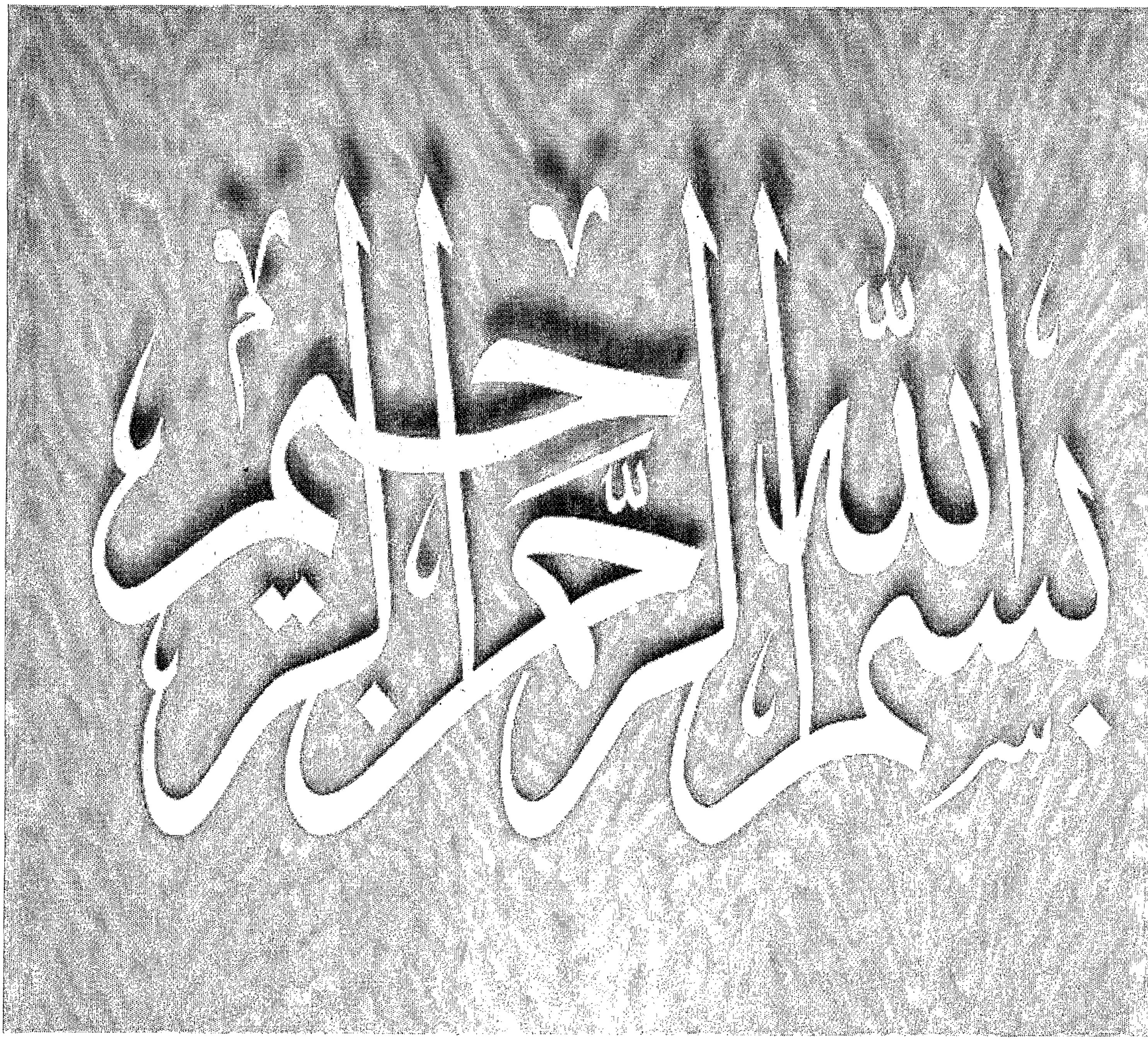
بسم الله الرحمن الرحيم

دانشگاه تربیت معلم

دانشکده علوم ریاضی و مهندسی کامپیوتر

پایان نامه کارشناسی ارشد آمار
(گرایش ریاضی)

عنوان:

ویژگی مقعر لوگی و ماکزیمم آنتروپی توزیع پواسون

استاد راهنما:

دکتر علی اکبر رحیم زاده ثانی

تدوین:

احسان قاسمی

بهمن ۱۳۸۶

بسمه تعالی

دانشکدهٔ علوم ریاضی و مهندسی کامپیوتر

تاریخ ....

شماره ....

پیوست ....

واحد ....

صورتجلسه دفاع از پایان نامه کارشناسی ارشد

جلسه دفاع از پایان نامه آقای **احسان قاسمی** دانشجوی دوره کارشناسی ارشد رشته ریاضی آمار تحت عنوان:

ویژگی مقعر لوگی و ماکزیمم آنتروپی توزیع پواسون

در روز سه شنبه مورخ ۸۶/۱۱/۳۰ در دانشکده علوم ریاضی و کامپیوتر تشکیل گردید و نتیجه آزمون به شرح زیر تعیین می گردد. نمره این آزمون ۱۸/۵ هجده و نیم می باشد.

- ۱- عالی ☑
- ۲- بسیار خوب ☐
- ۳- خوب ☐
- ۴- قابل قبول ☐
- ۵- غیرقابل قبول ☐

استاد راهنما	داور خارجی	داور داخلی
دکتر علی اکبررحیم زاده	دکتر حمید پزشک	دکتر عین اله پاشا

جواد لالی

رئیس دانشکده علوم ریاضی و کامپیوتر

تهران، خیابان طالقانی بین بهار وشریعتی، پلاک ۵۹۹ کدپستی ۱۵۶۱۸ تلفن۷۵۰۷۷۷۲ فاکس ۷۶۲۹۸۸

تقدیم به

روح **پدر** بزرگوارم که آرزوی دیدن چنین روزی را داشت

و

تقدیم به **مادرم** که هر چه دارم از دعاهای خالصانه اوست.

تقدیر و تشکر

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

درود و سلام می فرستم به روح بزرگ پدرم که نصایح و وصایای او راهنما و مشکل گشای من و خانواده ام در مسیر زندگی بوده و هست و امیدوارم که این هدیه ناچیز باعث شادی روحش گردد.

اعضای محترم خانواده ام خصوصاً مادرم، در تمامی مراحل زندگی یار و یاور من بوده اند و برای رسیدن من به این مرحله زحمات زیادی کشیده اند. از زحمات بی شائبه آنها که فراتر از آن چیزی بود که می توانست باشد، تشکر و قدردانی می کنم.

از جناب آقای دکتر علی اکبر رحیم زاده ثانی که در طول این دوره به من کمک های زیادی کردند و در تدوین این پایان نامه مرا یاری و راهنمایی نمودند سپاسگذارم.

در طول این دوره مطالب زیادی از جناب آقای دکتر عین الله پاشا فرا گرفتم. همچنین ایشان زحمت داروی این پایان نامه را بر عهده داشتند که بدین وسیله از ایشان تشکر و قدردانی می نمایم.

از جناب آقای دکتر حمید پزشک که به عنوان داور خارجی زحمت داوری این پایان نامه را تقبل کردند ممنون و سپاسگذارم.

از تمام دوستانی که در طول این دوره و نگارش این پایان نامه به من کمک های فراوانی کردند، مخصوصاً جناب آقایان اسرافیل رشیدی دیزجیکان، اسدا... یوسفی، حجت رنگین، حمیدرضا طاهری زاده زارج، شهاب بلوری، علی چاجی، مرتضی رئیسی، مسعود کلهر و مصطفی داوطلب علیایی تشکر و قدردانی می کنم.

احسان قاسمی

بهمن ۸۶

چکیده

در این پایان نامه، تحت قید ثابت بودن گشتاور اول، توزیع ماکزیمم آنتروپی را در دو خانواده خاص از متغیرهای تصادفی بدست می آوریم.

خانواده متغیرهای تصادفی دو جمله ای n-تعمیم یافته با میانگین λ را با $B_n(\lambda)$ نشان می دهیم و ثابت می کنیم توزیع دوجمله ای با پارامترهای $\left(n,\frac{\lambda}{n}\right)$ در خانواده $B_n(\lambda)$ دارای ماکزیمم آنتروپی است. این آنتروپی نسبت به n صعودی می باشد و وقتی n به ∞ میل می کند به صورت صعودی به آنتروپی توزیع پواسون با میانگین λهمگرا است. به علاوه اگر $ULC(\lambda)$ نشان دهنده خانواده متغیرهای تصادفی مقعر فرالوگی با میانگین λ باشد، نشان می دهیم که در این خانواده نیز توزیع پواسون با میانگین λ، دارای ماکزیمم آنتروپی است.

واژه های کلیدی : آنتروپی، توزیع پواسون، توزیع دو جمله ای، توزیع دوجمله ای تعمیم یافته، ماکزیمم آنتروپی، متغیر تصادفی مقعر فرالوگی، متغیر تصادفی مقعر لوگی.

فهرست مطالب

عنوان | صفحه

پیشگفتار

فصل اول . مقدمه ای بر آنتروپی

۱-۱ مقدمه ۲
۱-۲ میزان تعجب، عدم حتمیت و آنتروپی ۳
۱-۳ خواص آنتروپی ۱۱
۱-۴ آنتروپی توأم و شرطی متغیرهای تصادفی گسسته ۱۴
۱-۵ آنتروپی متغیرهای تصادفی پیوسته ۲۱
۱-۶ مدل ماکزیمم آنتروپی ۲۵
۱-۷ آنتروپی نسبی ۳۱
۱-۸ خواص آنتروپی نسبی ۳۲

فصل دوم. توزیع ماکزیمم آنتروپی در زیرمجموعه ای از خانواده متغیرهای تصادفی مقعر فرالوگی

۲-۱ مقدمه ۳۵
۲-۲ توزیع پواسون ۳۶
۲-۳ خاصیت مقعر لوگی ۳۷
۲-۴ همگرایی در آنتروپی نسبی ۴۸
۲-۵ کرانهایی برای آنتروپی نسبی یک متغیر تصادفی دوجمله‌ای تعمیم یافته ۵۰
۲-۶ آنتروپی در خانواده توزیعهای دوجمله‌ای تعمیم یافته ۵۹

فصل سوم. توزیع ماکزیمم آنتروپی در خانواده متغیرهای تصادفی مقعر فرالوگی

۳-۱ مقدمه ۶۶
۳-۲ نگاشتهای T_α، S_β و $V_{\alpha,\beta}$ ۶۷
۳-۳ نگاشت U_α و تابع $\Lambda(U_\alpha(X))$ ۷۹
۳-۴ تقعر آنتروپی در خانواده U_α و ماکزیمم بودن آنتروپی توزیع پواسون ۸۴

عنوان	صفحه

۵-۳ بحث و نتیجه گیری ۱۰۳

پیوست الف. واژه نامه فارسی به انگلیسی

پیوست ب. واژه نامه انگلیسی به فارسی

پیوست ج. مراجع

پیشگفتار

امروزه آنتروپی و نظریه اطلاع در رشته‌های آمار، مهندسی مکانیک، برق، نجوم، هواشناسی، اقتصاد، جغرافیا، تجارت بین‌الملل، بانکداری، پزشکی و... کاربرد دارد. آنتروپی از واژه یونانی Entropy به معنی «به درون خود می روم» گرفته شده است، این اصطلاح اولین بار توسط کلوسیوس[۱] [۳] در ترمودینامیک به کار گرفته شد. آنتروپی را می توان با بی‌نظمی معادل دانست. هر چه نظم سیستمی بالا رود آنتروپی آن کاهش می یابد و بالعکس کاهش نظم باعث افزایش آنتروپی می شود. در ترمودینامیک، آنتروپی اندازه بی‌نظمی در یک سیستم می باشد. سیستم از وضعیت آنتروپی پایین به سمت وضعیت آنتروپی بالا حرکت می کند.

بعد از کلوسیوس، بولتزمان[۲] [۳] از آنتروپی در تفسیرهای احتمالی در مبحثی از مکانیک آماری به شکل منسجم تری استفاده کرد. پلانک[۳] [۳] روابط بین آنتروپی و احتمال را گسترش داد تا اینکه سرانجام کلودشانون[۴] [۶] در سال ۱۹۴۸ برای اولین بار نظریه اطلاع را مطرح کرد.

مفهوم آنتروپی شانون هسته اصلی نظریه اطلاع را تشکیل می دهد. که گاهی اوقات تحت عنوان اندازه عدم قطعیت به آن اشاره می شود. این نظریه در آغاز با بسیاری از مسایل دشوار ریاضی روبرو بود، که قوانین اولیه شانون دقت کافی برای حل این مسائل را نداشت. مک میلان[5] و فین استین[۶] [۶] نظریه مفهومات اساسی منابع گسسته (منبع، کد، کانال و...) را به عنوان اولین تعاریف دقیق ریاضی به

[1] Clausius.
[2] Boltzman.
[3] Plank.
[4] Claude Shannon.
[5] Mc Millan.
[6] Feinstein.

کار بردند. کالبک[1] و لیبلر[2] [۱] در سال ۱۹۵۱ فاصله بین دو چگالی را تعریف کردند، که اطلاع تمییز دو چگالی را نمایش می دهد و آنتروپی نسبی نامیده می شود.

یکی از مهمترین کاربردهای آنتروپی در علم آمار، تعیین توزیع های مجهول با استفاده از اصل ماکزیمم آنتروپی است. این اصل در سال ۱۹۵۷ توسط جینس[3] [۶] بیان شد. در این روش با توجه به قیدهایی که در مورد گشتاورهای توزیع مجهول در نظر گرفته می شود، با ماکزیمم کردن آنتروپی، توزیع مجهول را تعیین می کنند.

در این پایان نامه پس از معرفی متغیرهای تصادفی مقعر لوگی و فرالوگی، توزیع ماکزیمم آنتروپی را در خانواده متغیرهای تصادفی مقعر فرالوگی، به شرط ثابت بودن گشتاور اول (میانگین) بدست می آوریم. بدین منظور ابتدا در فصل اول آنتروپی و ویژگی های آن را معرفی می کنیم، سپس در فصل دوم در زیرمجموعه ای از خانواده متغیرهای تصادفی مقعر فرالوگی به نام خانواده متغیرهای تصادفی دوجمله ای تعمیم یافته توزیع ماکزیمم آنتروپی را بدست می آوریم. در فصل سوم نیز توزیع ماکزیمم آنتروپی را در خانواده متغیرهای تصادفی مقعر فرالوگی بدست می آوریم. نشان می دهیم در هر دو خانواده مذکور توزیع پواسون دارای ماکزیمم آنتروپی است.

این پایان نامه بر اساس مقالات زیر تدوین شده است.

1) Johnson, O. Log-Concavity and the maximum entropy property of the poisson distribution, Stochastic processes and their applications. 117 (2007) 791-802.
2) Harremoës, P. Binomial and poisson distributions as maximum entropy distributions, IEEE Trans. Inform. Theory. 47 (5) (2001) 2039-2041.

---
[1] Kullback
[2] Leibler
[3] Jaynes

فصل اول

مقدمه ای برآنتروپی

۱-۱ مقدمه

برای نشان دادن اطلاع به صورت کمی، نظریه اطلاع به وجود آمد. همان طور که خصوصیات طول، مساحت، دما و... با یک عدد مشخص می شوند، میزان اطلاعی که هر موضوع به ما می دهد به وسیله نسبت تعداد سوالات لازم برای پی بردن به موضوع اندازه گیری می شود. این جواب‌های بدست آمده، که به صورت بله و خیر می باشد، را می توان با اعداد صفر و یک نشان داد. به همین دلیل واحد اطلاع را بیت می نامند.

اگر موضوعی که مورد نظر است، در فضایی غیر هم شانس قرار داشته باشد متوسط تعداد سؤالهایی که برای رسیدن به موضوع لازم است را 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. به عبارت دیگر پیام مربوط به حادثه ای که دارای کمترین احتمال وقوع است، بیشترین اطلاع را دارا است. همچنین می توان گفت $H(f)$ عبارت است از عدم حتمیت موجود در f برای قابل پیش بینی بودن برآمدی از X.

آنتروپی، یکنواختی یک توزیع را اندازه می گیرد. با افزایش $H(f)$ ، یکنواختی $f(x)$ بیشتر می شود. در نتیجه تراکم احتمال ها کمتر شده و پیش بینی برآمدی از X مشکل تر می‌شود.

اغلب آنتروپی را میزان تعجب یا میزان عدم حتمیت می نامند. به همین خاطر در ابتدای فصل به این موضوع می پردازیم.

در تدوین مطالب این فصل از [۲۱] استفاده شده است.

۱-۲ میزان تعجب، عدم حتمیت و آنتروپی

در طول زندگی، ما به مسائل و اتفاقاتی بر می خوریم که نتیجه بعضی از آنها قطعی و حتمی است. و برخی دیگر نامشخص است. مثلا، اگر در باک بنزین ماشین، بنزین وجود نداشته باشد، ماشین روشن نخواهد شد. در این حالت نتیجه آزمایش (تلاش برای روشن کردن ماشین بدون بنزین) معلوم و حتمی است و ما قبل از انجام آزمایش، می دانیم که ماشین روشن نخواهد شد و نتیجه آزمایش باعث تعجب ما نخواهد شد.

حال فرض کنید اخبار هواشناسی خبر بارش برف در شهر اهواز را اعلام نماید. مطمئنا این خبر باعث تعجب شما خواهد شد. چرا که انتظار شنیدن چنین خبری را نداشته‌اید. اما خبر بارش برف در شهر اردبیل برای شما خبری عادی است. شما با شنیدن این خبر خیلی کمتر از شنیدن خبر بارش برف در اهواز تعجب خواهید کرد (شاید اصلا تعجب نکنید.)

اگر یک سکه نااریب را پرتاب کنیم، احتمال وقوع شیر با احتمال وقوع خط برابر و برابر $\frac{۱}{۲}$ است.

$$P(H)=P(T)=\frac{۱}{۲}, \quad (\text{H:شیر و T:خط}).$$

حال سکه دیگری را در نظر بگیرید که اریب باشد. مثلا یک طرف آن برآمده باشد. در این حالت احتمال وقوع شیر با احتمال وقوع خط برابر نخواهد بود. فرض کنید احتمال آمدن شیر ۰/۸ و احتمال وقوع خط ۰/۲ باشد. در این حالت با اطمینان بیشتری نسبت به حالت اول (سکه نا اریب) می توان وقوع شیر را در هر بار پرتاب سکه پیش بینی کرد. پس این حالت دارای درجه بلاتکلیفی کمتری نسبت به حالت اول است. دیدن شیر در پرتاب سکه اریب (حالت دوم) کمتر باعث تعجب ما می شود. چرا که برآمدی اتفاق افتاده که احتمال رخ دادن آن بیشتر است. اما دیدن خط در پرتاب این سکه باعث

تعجب ما خواهد شد. پس به طور کلی رخ دادن پیشامدی که حتمیت کمتری دارد، باعث تعجب بیشتر ما خواهد شد.

عدم حتمیت سختی پیش بینی برآمدهای متغیر تصادفی X را تشریح می کند. فرض کنید متغیر تصادفی X مقادیر $x_1, x_2, \dots, x_n$ را با احتمال‌های $p_1, p_2, \dots, p_n$ اختیار کند. ما به دنبال یافتن کمیتی هستیم که عدم حتمیت X را اندازه گیری نماید. عدم حتمیت را که تابعی از $p_1, p_2, \dots, p_n$ است با $H(p_1, p_2, \dots, p_n)$ یا به اختصار با $H(X)$ نشان می دهیم. $H(p_1, p_2, \dots, p_n)$ را می توان به صورت میانگین وزنی اعداد $h(p_1), h(p_2), \dots, h(p_n)$ در نظر گرفت که $h(p_i)$ عدم حتمیت برآمد $\{X = x_i\}$ یا به عبارت دیگر میزان تعجب حاصل از اطلاع وقوع برآمد $\{X = x_i\}$است. تابع $H(X)$ متوسط عدم حتمیت برآمدهای $\{X = x_i\}$ است.

برای بدست آوردن شکل تابعی H، شرایطی را که از نظر شهودی منطقی به نظر می رسند بر H تحمیل می کنیم. ابتدا فرض کنید که مقادیر X هم شانس باشند. متوسط عدم حتمیت m برآمد هم‌شانس را با $L(m)$ نشان می دهیم، یعنی $L(m) = H\left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right)$. برای مثال $L(2)$ عدم حتمیت مرتبط با پرتاب یک سکه نا اریب است، در حالیکه $L(6)$ عدم حتمیت مرتبط با پرتاب یک تاس است. انتظار داریم که عدم حتمیت حالت دوم بزرگتر از اولی باشد. در واقع، اولین شرط ما روی عدم حتمیت این است که:

$L(m) = H\left(\frac{1}{m}, \frac{1}{m}, \dots, \frac{1}{m}\right)$ تابع یکنوای صعودی از m است. یعنی اگر $m < m'$ آنگاه:

$$L(m) < L(m') \;,\; (m, m' = 1, 2, \dots).$$

حال آزمایشی را با دو متغیر تصادفی مستقل X و Y در نظر بگیرید. فرض کنید X مقادیر $x_1, x_2, \ldots, x_m$ را با احتمال‌های یکسان و Y مقادیر $y_1, y_2, \ldots, y_l$ را با احتمال‌های یکسان اختیار کند. بنابراین آزمایش توام شامل X و Y دارای ml برآمد هم شانس است و لذا عدم حتمیت آزمایش توام برابر با $L(ml)$ است. اگر مقدار X مشخص شود، متوسط عدم حتمیت Y به دلیل فرض استقلال نباید تغییر کند. بنابراین انتظار داریم که عدم حتمیت توام X و Y، منهای متوسط عدم حتمیتی که در نتیجه مشخص شدن مقدار X از میان می رود، برابر متوسط عدم حتمیت Y باشد. لذا شرط دوم روی اندازه عدم حتمیت این است که:

$$L(ml) = L(m) + L(l) \ , \ (m, l = 1, 2, \ldots) .$$

حال قید هم شانس بودن برآمدها را حذف کرده، به حالت کلی بر می گردیم. مقادیر متغیر تصادفی X را به دو گروه A و B تقسیم می کنیم، که A شامل $x_1, x_2, \ldots, x_r$ و B شامل $x_{r+1}, x_{r+2}, \ldots, x_m$ است. آزمایش مرکبی را به صورت زیر می سازیم.

ابتدا یکی از دو گروه را انتخاب می کنیم. گروه A با احتمال $p_1 + p_2 + \ldots + p_r$ و گروه B با احتمال $p_{r+1} + p_{r+2} + \ldots + p_m$ انتخاب می شود. اگر گروه A انتخاب شود x_i را با احتمال $\frac{p_i}{\sum_{i=1}^{r} p_i}$ $(i = 1, 2, \ldots, r)$ انتخاب می‌کنیم که احتمال شرطی x_i به شرط بودن مقادیر X در گروه A است. به همین ترتیب اگر گروه B انتخاب شود، x_i را با احتمال $\frac{p_i}{\sum_{i=r+1}^{m} p_i}$ $(i = r+1, r+2, \ldots, m)$ انتخاب می کنیم. آزمایش مرکب معادل آزمایش اولیه است. زیرا اگر Y نتیجه آزمایش مرکب باشد، احتمال اینکه $Y = x_1$ باشد برابر است با

$$P(Y=x_1)=P(\text{انتخاب شود و } x_1 \text{ انتخاب گردد } A)$$

$$=P(\text{انتخاب شود } A)P(\text{انتخاب شود} | x_1 \text{ انتخاب شود } A)=$$

$$=\sum_{i=1}^{r}p_i\left(\frac{p_1}{\sum_{i=1}^{r}p_i}\right)=p_1.$$

به همین ترتیب، $P(Y=x_i)=p_i$ $(i=1,2,\dots,m)$. لذا Y و X توزیع یکسانی دارند. قبل از اینکه آزمایش مرکب انجام شود، متوسط عدم حتمیت $H(p_1,p_2,\dots,p_m)$ است. اگر مشخص شود که کدام یک از دو گروه A یا B انتخاب شده است، به طور متوسط عدم حتمیتی به اندازه $H(p_1+p_2+\dots+p_r,p_{r+1}+p_{r+2}+\dots+p_m)$ از بین می رود. گروه A با احتمال $p_1+p_2+\dots+p_r$ انتخاب می شود و عدم حتمیت باقیمانده عبارت است از:

$$H\left(\frac{p_1}{\sum_{i=1}^{r}p_i},\frac{p_2}{\sum_{i=1}^{r}p_i},\dots,\frac{p_r}{\sum_{i=1}^{r}p_i}\right).$$

گروه B با احتمال $p_{r+1}+p_{r+2}+\dots+p_m$ انتخاب می شود و عدم حتمیت باقیمانده عبارت است از:

$$H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{m}p_i},\frac{p_{r+2}}{\sum_{i=r+1}^{m}p_i},\dots,\frac{p_m}{\sum_{i=r+1}^{m}p_i}\right).$$

بنابراین، متوسط عدم حتمیت باقیمانده پس از انتخاب گروه ها عبارت است از:

$$(p_1+\dots+p_r)H\left(\frac{p_1}{\sum_{i=1}^{r}p_i},\dots,\frac{p_r}{\sum_{i=1}^{r}p_i}\right)+(p_{r+1}+\dots+p_m)H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{m}p_i},\dots,\frac{p_m}{\sum_{i=r+1}^{m}p_i}\right).$$

انتظار داریم، متوسط عدم حتمیت آزمایش مرکب منهای متوسط عدم حتمیتی که با مشخص شدن یک گروه، از میان می رود، برابر متوسط عدم حتمیتی باشد که پس از مشخص شدن این گروه باقی می ماند. لذا شرط سوم عبارت است از:

$$\begin{aligned}H(p_1,p_2,\dots,p_m)&=H(p_1+p_2+\dots+p_r,p_{r+1}+p_{r+2}+\dots+p_m)\\&+(p_1+p_2+\dots+p_r)H\left(\frac{p_1}{\sum_{i=1}^{r}p_i},\frac{p_2}{\sum_{i=1}^{r}p_i},\dots,\frac{p_r}{\sum_{i=1}^{r}p_i}\right)\\&+(p_{r+1}+p_{r+2}+\dots+p_m)H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{m}p_i},\frac{p_{r+2}}{\sum_{i=r+1}^{m}p_i},\dots,\frac{p_m}{\sum_{i=r+1}^{m}p_i}\right).\end{aligned}$$

سرانجام یک خصوصیت ریاضی بر H تحمیل می کنیم و آن اینکه $H(p,1-p)$ تابعی پیوسته از p می باشد. به عبارت دیگر، انتظار داریم که هر تغییر کوچکی در احتمال مقادیر X باعث تغییری کوچک در عدم حتمیت X گردد.

به طور خلاصه، چهار شرط زیر را به عنوان اصول فرض می کنیم:

۱. $L(m)=H\left(\frac{1}{m},\frac{1}{m},\dots,\frac{1}{m}\right)$ تابعی صعودی از m است . $(m=1,2,\dots)$.

۲. $L(ml)=L(m)+L(l)$ $(m,l=1,2,\dots)$.

۳. $$H(p_1,p_2,\dots,p_m)=H(p_1+p_2+\dots+p_r,p_{r+1}+p_{r+2}+\dots+p_m)$$
$$+(p_1+p_2+\dots+p_r)H\left(\frac{p_1}{\sum_{i=1}^{r}p_i},\frac{p_2}{\sum_{i=1}^{r}p_i},\dots,\frac{p_r}{\sum_{i=1}^{r}p_i}\right)$$
$$+(p_{r+1}+p_{r+2}+\dots+p_m)H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{m}p_i},\frac{p_{r+2}}{\sum_{i=r+1}^{m}p_i},\dots,\frac{p_m}{\sum_{i=r+1}^{m}p_i}\right).$$

(اصل ۳ اصل گروه بندی نامیده می شود.)

۴. $H(p,1-p)$ تابعی پیوسته از p می باشد.

چهار اصل فوق، اندازه عدم حتمیت را به طور کامل مشخص می کنند.

قضیه ۱-۱ تنها تابعی که در چهار اصل فوق صدق می کند به صورت زیر است:

$$H(p_1,p_2,\dots,p_n)=-C\sum_{i=1}^{n}p_i\log p_i\ . \qquad (1-1)$$

که در آن C عدد مثبت دلخواهی است و پایه لگاریتم نیز عددی بزرگتر از یک است.

اثبات . صفحه ۸ از مرجع شماره [۱].

چون $\log_a^x=\log_a^b\log_b^x$ لذا تغییر پایه لگاریتم در فرمول H، معادل تغییر ثابت C و یا به عبارت دیگر معادل تغییر واحد عدم حتمیت است. در عمل، C را برابر ۱ فرض می کنند. معمولا در حالت گسسته پایه لگاریتم را ۲ اختیار می کنند. در این حالت واحد عدم حتمیت، بیت (bit)، مخفف واژه Binary Digits (ارقام دو دویی) است.

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

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

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

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 $H(X)$ می توان از نماد $H(f(X))$ یا $H(f)$ یا $H(p_۱, p_۲, ...)$ 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 صرفا تغییر در مقیاس واحدها می باشد. اما اغلب پایه را ۲ یا e (عدد نپر) قرار می دهند. واحد آنتروپی را هنگامی که در پایه ۲ محاسبه می‌شود bit می نامند و هنگامی که در پایه e محاسبه شود، واحد اطلاع بر حسب nat (واحد طبیعی) بیان می شود

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

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

$$= -(۱-p)\log(۱-p) - p\log p$$

$$= \log \frac{(۱-p)^{p-۱}}{p^p} .$$

مثال ۱-۲ . آنتروپی متغیر تصادفی هندسی X با پارامتر p به صورت زیر بدست می آید.

$$H(X) = -\sum_{x=0}^{\infty} p(1-p)^x \log\left(p(1-p)^x\right)$$

$$= -\sum_{x=0}^{\infty} p(1-p)^x \left(\log p + x\log(1-p)\right)$$

$$= -p\log p\sum_{x=0}^{\infty}(1-p)^x - p\log(1-p)\sum_{x=0}^{\infty}x(1-p)^x$$

$$= -\log p - \frac{1-p}{p}\log(1-p)$$

$$= \frac{\log\frac{(1-p)^{p-1}}{p^p}}{p} = \frac{h(p)}{p}.$$

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

اگر در $H(p_1, p_2, \ldots)$ احتمال‌ها را به ترتیب متفاوتی قرار دهیم، مقدار آنتروپی تغییر نمی کند. اما طبق مثال زیر توزیع‌های احتمال متفاوت، ممکن است دارای آنتروپی یکسانی باشند.

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

$$P = (0/5, 0/25, 0/25).$$

$$Q = (0/48, 0/32, 0/2).$$

در اینصورت:

$$H(P) = H(Q) = 1.5\,\text{bit}.$$

پس $H(X)$ تابع توزیع را به طور یکتا مشخص نمی کند.