دانشکده علوم پایه و مهندسی کامپیوتر

(گرایش ریاضی)

عنوان :

ساختار مقایسه‌ای برآوردگرهای جایگذاری نرخ آنتروپی
در
زنجیر مارکف دوحالتی

استاد راهنما:

دکتر عین اللّه پاشا

تدوین:

یاسر چمانی

فهرست مطالب

پیشگفتار

زنجیرهای مارکف و آنتروپی از زمانی که شانون [۱] در سال ۱۹۴۸ آنتروپی را در نظریه احتمال مطرح کرد، به هم پیوند یافتند.
شانون آنتروپی توزیع P با فضای حالت متناهی E را به صورت زیر بیان کرد

$$S(P) = -\sum_{i\in E} P(i)\log P(i)$$

با این قرارداد که $0 \log 0 = 0$ [۳۶]
برای یک فرایند تصادفی زمان گسسته $X = \{X_n : n \geq 0\}$، گیراردین [۲] آنتروپی در زمان n را به عنوان آنتروپی شانون توزیع حاشیه‌ای n—بعدی X به صورت زیر تعریف کرد

$$H_n(X) = -\sum_{i\in E} P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)\log P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)$$

تحت شرایط مناسب $\frac{1}{n}H_n(X)$ وقتی که n به بینهایت میل می‌کند، همگراست. زمانی که $\lim_{n\to\infty}\frac{1}{n}H_n(X)$ موجود باشد این آنتروپی را نرخ آنتروپی فرآیند می‌نامند[۲۰].
نرخ آنتروپی عدم قطعیت را در یک سیستم دینامیکی مشخص می‌کند، مسیو و جاناتان شلنس [۳] نیز نرخ آنتروپی را حد مجانبی یک سری زمانی تعریف می‌کنند[۲۶].
نرخ آنتروپی برای اولین بار توسط شانون برای یک زنجیر مارکف همگن ارگودیک با فضای حالت متناهی E، به عنوان مجموع آنتروپی‌های احتمالات انتقال $(P(i,j))_{i,j\in E}$، که احتمال رخداد هر حالت، طبق توزیع مانای π زنجیر، وزن‌دار شده‌اند به صورت زیر تعریف شد

$$H(X) = -\sum_{i\in E}\pi(i)S(p_i)$$

[۱] Shannon
[۲] Girardin
[۳] Matthew and jonathon Shlens

که $P_i = (P(i,j))_{j \in E}$ لذا

$$H(X) = -\sum_{i \in E} \pi(i) \sum_{j \in E} P(i,j) \log P(i,j)$$

شانون در سال ۱۹۴۸ همگرایی $\frac{1}{n} \log P(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n)$ را به $H(X)$ در احتمال ثابت کرد[۳۶]. امسی میلان [۴] در سال ۱۹۵۳ همگرایی در میانگین برای هر فرایند ارگودیک مانا با فضای حالت متناهی را ثابت کرد[۲۷].
در نظریه اطلاع، نرخ آنتروپی منبع، اندازه‌ای از میزان پیچیدگی (بی‌نظمی) منبع است. اما نرخ آنتروپی در بیشتر زمینه‌های کاربردی دیگر نیز مورد استفاده قرار می‌گیرد. به عنوان مثال در کدگذاری و پیچیدگی الگوریتم‌ها[۲۱].
زمانی که مشاهداتی از فرایند موجود باشد نیاز به برآورد کردن نرخ آنتروپی بوضوح پیدا می‌شود. تقریب‌هایی از آنتروپی را می‌توان به وسیله الگوریتم‌های عددی بدست آورد. الگوریتم زیو—لمپل [۵] برای تقریب آنتروپی در یک فرایند باینری به کار می‌رود. پلوتکن و واینر[۶] در سال ۱۹۶۵ یک برآورد کننده الگوریتمی از نرخ آنتروپی را برای یک مسئله صف‌بندی در شبکه ارتباط از راه دور بیان کردند[۳۱]. بت [۷] در سال ۱۹۵۹ برآوردی برای توابع دقیقی از ماتریس انتقال نشان داد که ممکن است برای نرخ آنتروپی به کار رود[۱۵].
میسو یجیوس [۸] در سال ۱۹۶۶ برآوردگر جایگذاری نرخ آنتروپی را برای زنجیر مارکف ارگودیک مانا با فضای حالت متناهی مطرح کرد[۲۸].
موخامدخانوا[۹] در سال ۱۹۸۴ ویژگی‌های مجانبی متفاوتی از برآوردگر جایگذاری، برای زنجیر ارگودیک مانای دوحالتی تعیین کرد که در اصطلاح، سری‌های طرح مشاهده شده می‌نامند[۳۰].
گیراردین و کیوپریا[۱۰] در سال ۲۰۰۷ ثابت کردند که برای زنجیر مارکف ارگودیک شمارا، برآوردگرهای جایگذاری نرخ آنتروپی سازگارند. همچنین برای فضای حالت متناهی و احتمالات انتقال غیریکنواخت نرمال مجانبی بودن حفظ می‌شود ولی در حالت کلی نمی‌توان یک عبارت دقیقی برای واریانس جامعه معلوم باشد[۱۶].
اگر از یک مسیر بلند، چندین نمونه مجزا کوتاه گرفته شود مدل از چندین مسیر مناسب است. عملاً مشاهده تعداد زیادی مسیرهای مستقل از زنجیر با طول کوتاه نسبت به یک مسیر با طول بلند آسانتر است. مخصوصاً در کاربردهای داده‌های بقا(اندرسن و گودمن [۱۱]) [۹].
پایان نامه‌ای که پیش رو دارید شامل ۴ فصل است که در فصل اول مقدماتی از آنتروپی و در

[۴] Mc Millan
[۵] Ziv-Lempel
[۶] Plothin & Vyner
[۷] Bhat
[۸] Misevichyus
[۹] Mukhamedkhanova
[۱۰] Girardin & Ciuperea
[۱۱] Anderson & Goodman

فصل دوم در مورد همگرایی دنباله‌ها و برآوردگر جایگذاری و روش دلتا و در فصل سوم در مورد فرایندهای تصادفی و استنباط آماری در مورد زنجیر مارکف و نرخ آنتروپی فرایند تصادفی صحبت می‌کنیم.
در فصل چهارم که شالوده اصلی این پایان نامه است زنجیر مارکف دوحالتی با فضای حالت $E = \{۰, ۱\}$ را در نظر می‌گیریم.
زنجیرهای مارکف که مقادیر را به یک فضای دوحالتی می‌برند یک ابزار مفید در ساختن تعداد زیادی موقعیت‌های حقیقی را تشکیل می‌دهند. آنها در بسیاری زمینه‌های کاربردی شامل شبکه ارتباطات از راه دور، قابلیت اعتماد(پایایی)، تحلیل بقا و هواشناسی ظاهر می‌شوند.
در زنجیر مارکف دوحالتی، توزیع مانا را می‌توان به راحتی و به صورت تحلیلی از احتمالات انتقال بدست آورد. نرخ آنتروپی که تابع دقیقی از احتمالات انتقال بوده شامل دو پارامتر است.
برآوردی از احتمالات انتقال را بر اساس مشاهداتی از یک مسیر با طول بلند یا چندین مسیر با طول ثابت را با ۳ روش تجربی، MLE و ترکیبی بدست آورده و با جایگذاری این برآوردگرها به جای پارامترهایشان در نرخ آنتروپی، برآوردگر جایگذاری نرخ آنتروپی را بدست می‌آوریم.
احتمالات انتقال به دو صورت یکنواخت و غیریکنواخت است. توزیع مجانبی برآوردگرهای جایگذاری را که از هر یک از این احتمالات انتقال بدست می‌آید را بررسی می‌کنیم و هر یک از این برآوردگرهای جایگذاری را به طور تحلیلی یا به واسطه شبیه سازی مقایسه می‌کنیم.
این پایان نامه بر اساس مقالات زیر تدوین شده است

[1]Sesboue A, Girardin V(2008) comparative construction of Plug-in estimator of the entropy rate of two-state markov chains.Springer Science+Business media,LLC

[2]Ciupercac.Girardin V(2007)estimation of the entropy rate of a countable markov chain. commun stat theory methods 36(14):2493-2508

چکیده:

در این پایان نامه ساختار مقایسه‌ای برآوردگرهای جایگذاری نرخ آنتروپی در زنجیرهای مارکف دوحالتی را مطالعه می‌کنیم. نرخ آنتروپی زنجیر مارکف همگن ارگودیک، تنها دو مقداری را می‌گیرد که تابع دقیقی از احتمالات انتقالش بوده و شامل دو پارامتر است. برآوردی از احتمالات انتقال را با سه روش (تجربی، MLE، ترکیبی) بر اساس مشاهداتی از یک مسیر با طول بلند و یا چندین مسیر با طول ثابت را بدست آورده و با جایگذاری این برآوردگرها به جای پارامترهایشان در نرخ آنتروپی، برآوردگر جایگذاری نرخ آنتروپی را بدست می‌آوریم و همچنین توزیع مجانبی برآوردگرها را بررسی می‌کنیم.
در پایان بررسی عددی جزئیات، با استفاده از نتایج شبیه‌سازی فراهم شده و داده‌های واقعی از رخداد بارندگی استخراج شده است.

واژگان کلیدی: نرخ آنتروپی، زنجیر مارکف همگن دوحالتی، برآوردگر جایگذاری.

فصل ۱

آنتروپی

۱—۱ مقدمه

امروزه آنتروپی، نظریه‌ی اطلاع در رشته های آمار، مهندسی مکانیک، برق، نجوم، اقتصاد، پزشکی، مدیریت و سایر رشته‌ها کاربرد دارد.
آنتروپی از واژه یونانی *Entropy* به معنی « به درون خود می‌روم» گرفته شده است، این اصطلاح اولین بار توسط کلوزیوس [۱] [۱۷] در ترمودینامیک به کار گرفته شد.
پایه و اساس مفهوم آنتروپی در سال ۱۸۷۷ توسط بولتزمان [۲] مطرح شد[۱۷] و در سال ۱۹۴۸ بوسیله‌ی شانون [۳] و قبل از وی افراد دیگر همچون اچ . نیکوست [۴] در تفسیر آنتروپی از نقطه نظر تئوری اطلاع مطالعاتی انجام داده اند.
نظریه‌ی اطلاع برای نخستین بار توسط شانون در سال ۱۹۴۸ مطرح شد. در واقع می‌توان گفت: که مفهوم آنتروپی شانون هسته اصلی نظریه اطلاع را تشکیل می‌دهد، که گاهی اوقات تحت عنوان اندازه عدم قطعیت به آن مراجعه می‌کنیم. آنتروپی موضوعی است که به نوعی به عدم قطعیت و اطلاع مربوط بوده و در رابطه ترمودینامیک آماری و رمز گذاری بکار می رود.
آنتروپی را می توان با بی‌نظمی معادل دانست. هر چه نظم سیستمی بالا رود آنتروپی آن کاهش می‌یابد و بالعکس کاهش نظم باعث افزایش آنتروپی می‌شود. آنتروپی یک سیستم با میزان

---
[۱] Clausius
[۲] Ludwing Boltzman
[۳] Shannon
[۴] H.Nyquist

اطلاع موجود در آن مرتبط است، یک سیستم با نظم بیشتر می‌تواند با بایت های کمتری از اطلاع توصیف شود در حالیکه سیستمی با بی‌نظمی بالاتر برای توصیف شدن به بایت های بیشتری از اطلاعات نیازمند است. نظریه‌ی اطلاع در ابتدا برای بیان عددی اطلاع بوجود آمد. همانطور که فاصله، دما، زمان، ... را با عدد اندازه گیری می کنند، مقدار اطلاعی را که این موضوع به ما می دهد، بوسیله تعداد سوالات مثبت لازم برای پی بردن به آن موضوع اندازه گیری می‌شود. این جوابهای بدست آمده را که بصورت بله و خیر می باشد، می‌توان با اعداد صفر و یک نشان داد، به همین دلیل است که واحد اطلاع را بیت (bit) می‌نامیم. بعنوان مثال 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. بصورت شهودی مقدار اطلاعات دریافتی از وقوع یک حادثه با احتمال وقوع این حادثه نسبت عکس دارد. به عبارت دیگر پیام مربوط به حادثه‌ای که دارای کمترین احتمال وقوع است، بیشترین اطلاع را داراست. اغلب آنتروپی را میزان تعجب یا میزان عدم حتمیت می‌نامند.

۱—۲ آنتروپی (میزان تعجب ، عدم قطعیت)

پیشامد A را در نظر بگیرید که می تواند رخ دهد ، یعنی احتمال وقوع آن صفر نباشد، اگر آزمایش انجام گیرد و گفته شود که A رخ داده است چقدر تعجب می‌کنیم؟ به نظر منطقی است که تصور کنیم میزان تعجب حاصل از اطلاع وقوع A تنها باید به احتمال وقوع A بستگی داشته باشد. برای مثال فرض کنیم $P(A) = ۰/۹۹۹$، در اینصورت حتم داریم که A اتفاق می‌افتد. حال اگر $P(A) = ۰/۰۰۱$، بطور منطقی A یا رخ نمی‌دهد یا احتمال رخ دادن آن خیلی کمتر است.

در این قسمت سعی می‌کنیم میزان تعجب را کمی کنیم. در آغاز روی این اصل توافق می‌کنیم که میزان تعجب از وقوع یک پیشامد A فقط بستگی به احتمال وقوع آن دارد و آن را با $h(p)$ نشان می‌دهیم، یعنی میزان تعجبی که از وقوع پیشامدی با احتمال p حاصل می‌گردد. حال سعی می‌کنیم شکل تابعی $h(p)$ را براساس توافق روی مجموعه‌ای از شرایط منطقی تعیین نماییم. همچنین فرض می‌کنیم که $h(p)$ برای مقادیر $۰ < p \leq ۱$ تعریف شده و برای پیشامدی که $p = ۰$، $h(p)$ تعریف نشده است.

شرط اول، بیان این واقعیت است که وقتی بشنویم پیشامد حتمی که باید اتفاق بیفتد، رخ دهد، تعجب از وقوع آن صفر است.

اصل ۱

$$h(۱) = ۰$$

شرط دوم، بیان این واقعیت است که میزان تعجب ما از وقوع پیشامدی که شانس کمی برای رخ دادن دارد بیشتر از تعجب ما برای پیشامدی است که شانس بیشتری برای رخ دادن دارد.

اصل ۲ $h(p)$ 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 می رود هر تغییر کوچک در p باعث تغییر کوچک در $h(p)$ می شود.

اصل ۳ $h(p)$ تابعی پیوسته از p است.

شرط چهارم، دو پیشامد مستقل A و B را در نظر می گیریم به طوری که $P(A) = p$ و $P(B) = q$ ، چون احتمال وقوع توام برابر $P(AB) = pq$ است، پس میزان تعجب حاصل از وقوع توام A و B برابر $h(pq)$ است. حال فرض می 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، بنابراین $h(pq) - h(p)$ نشان دهنده افزایش تعجب است، وقتی مطلع شویم که B نیز رخ داده است. به علاوه چون A و B مستقل از هم هستند و اطلاع از وقوع یا عدم وقوع پیشامد A تأثیری در احتمال وقوع B ندارد، بنابراین افزایش تعجب بایستی دقیقاً برابر با $h(p)$ باشد.

اصل ۴ به ازای $۰ < p \leq ۱$ ، $۰ < q \leq ۱$

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

چهار اصل فوق عدم حتمیت را به طور کامل مشخص می کنند.
اکنون قضیه زیر را که بیان کننده ضابطه تابع $h(p)$ است مطرح می کنیم.

قضیه ۱.۱ تابع $h(p)$ در اصول ۱ تا ۴ صدق می کند، اگر و فقط اگر $h(p) = -c\log p$ که c یک عدد دلخواه صحیح و مثبت و مبنای لگاریتم همواره بزرگتر از یک است.
اثبات در مرجع [٦] آمده است.

۱—۳ آنتروپی متغیرهای تصادفی گسسته

متغیر تصادفی X که مقادیر $x_۱, x_۲, \ldots, x_n$ را با احتمالات متناظر $p_۱, p_۲, \ldots, p_n$ اختیار می کند را در نظر می گیریم. چون $-\log p_i$ نشان دهنده‌ی میزان تعجب حاصل از این است که متغیر تصادفی X مقدار x_i را اختیار می کند، در نتیجه میانگین موزون (امید ریاضی) میزان تعجب حاصل از این اطلاع در مورد مقدار متغیر تصادفی X برابر است با

$$H(X) = -\sum_{i=۱}^{n} P(x_i)\log P(x_i)$$

که $P(x_i) = P(X = x_i)$

کمیت $H(X)$ در نظریه اطلاع با عنوان آنتروپی متغیر تصادفی X شناخته می شود. چون ممکن است $P(X = x_i) = ۰$ باشد لذا $P(X = x_i) \log P(X = x_i)$ می تواند نامعین باشد. بنابراین هر گاه $P(X = x_i) = ۰$ بود، طبق قرارداد $P(X = x_i) \log P(X = x_i)$ را برابر صفر در نظر می گیریم.

تعریف ۲.۱ فرض کنیم X یک متغیر تصادفی گسسته با تکیه گاه S_X و تابع جرم احتمال $f(x)$ باشد، در این صورت آنتروپی X به صورت زیر تعریف می شود

$$H(X) = -E(\log f(X)) = -\sum_{x \in S_X} f(x) \log f(x)$$

یاد آوری 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 $H(f(X))$ و یا $H(f)$ و یا $H(p_۱, p_۲, \ldots)$ استفاده کرد.

در تعریف (۲.۱)، پایه لگاریتم تعریف نشده است. در اکثر موارد مهم نیست که آن را چه عددی قرار می دهیم ، چون $\log_a(x) = \log_a b . \log_b x$؛ لذا تغییر پایه لگاریتم در فرمول $H(X)$ معادل تغییر ثابت c ویا به عبارت دیگر معادل تغییر واحد عدم حتمیت است. در عمل همانطور که گفته شد c را برابر یک فرض می کنند و معمولاً در حالت گسسته پایه لگاریتم را ۲ یا e انتخاب می کنند.

واحد آنتروپی را هنگامی که در پایه ۲ محاسبه می شود bit (مخفف اعداد دودویی) می نامند و هنگامی که در پایه e محاسبه می شود واحد اطلاع بر حسب nat (واحد طبیعی) بیان می شود. رابطه ی بین bit و nat به صورت زیر است

$$۱ bit = ۰/۶۹۳ nat$$

مثال ۳.۱ آنتروپی متغیر تصادفی هندسی X با پارامتر p به صورت زیر است

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

لم ۴.۱ برای هر $x > \circ$
داریم $\ln x \leq x - ۱$ و تساوی زمانی برقرار است که $x = ۱$ باشد.
اثبات در مرجع [۸] آمده است.

قضیه ۵.۱ فرض کنیم $S_X = \{x_۱, \dots, x_n\}$، تکیه‌گاه متغیر تصادفی X باشد و مولفه‌های بردار احتمال $P = \{p_۱, \dots, p_n\}$ به ترتیب احتمال‌های وابسته به پیشامدهای $\{X = x_i\}$ باشند، در این صورت

الف) آنتروپی X کمیتی نامنفی است، یعنی $H(X) \geq \circ$
ب) آنتروپی X صفر است اگر و تنها اگر X مقداری ثابت را با احتمال یک بگیرد.
ج) حداکثر آنتروپی هنگامی است که X دارای توزیع یکنواخت باشد. به به عبارت دیگر $H(X) \leq \log n$ و تساوی برقرار است اگر و تنها اگر

$$p_۱ = p_۲ = \ldots = p_n = \frac{۱}{n}$$

اثبات در مرجع [۱۷] آمده است.

تذکر ۶.۱ برای بررسی ویژگیهای آنتروپی متغیّرهای تصادفی گسسته به مرجع [۲۳] رجوع شود.

۱—۳—۱ آنتروپی متغیرهای تصادفی توأم

دو متغیر X و Y را که به ترتیب مقادیر $x_۱, x_۲, \ldots, x_n$؛ $y_۱, y_۲, \ldots, y_n$ را با تابع احتمال توأم زیر اختیار می‌کنند، در نظر می‌گیریم

$$P(x_i, y_i) = P(X = x_i, Y = y_i)$$

آنگاه میزان عدم قطعیت یا آنتروپی توأم برای مقدار بردار تصادفی (X, Y) که با $H(X, Y)$ نشان داده می‌شود برابر است با

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

نتیجه ۷.۱ اگر n متغیر تصادفی $X_۱, X_۲, \ldots, X_n$ مقادیر $x_{i۱}, x_{i۲}, \ldots, x_{in}$ را با تابع احتمال توأم زیر اختیار کنند

$$P(x_{i۱}, x_{i۲}, \ldots, x_{in}) = P(X_۱ = x_{i۱}, X_۲ = x_{i۲}, \ldots, X_n = x_{in})$$

آنگاه آنتروپی توأم بردار تصادفی $(X_1, X_2, \ldots, X_n)$ را با $H(X_1, X_2, \ldots, X_n)$ نشان می‌دهیم و برابر است با

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= -E(\log P(X_1, X_2, \ldots, X_n)) \\ &= -\sum_{x_{i1}} \sum_{x_{i2}} \ldots \sum_{x_{in}} P(x_{i1}, \ldots, x_{in}) \log P(x_{i1}, \ldots, x_{in}) \end{aligned}$$

۱–۳–۲ آنتروپی شرطی

فرض کنیم که مقدار $Y = y_i$ مشاهده شده باشد، در این حالت مقدار عدم قطعیت باقی‌مانده X برابر است با

$$H(X|Y = y_i) = -\sum_i P(x_i|y_i) \log P(x_i|y_i)$$

به طوری که

$$P(x_i|y_i) = P(X = x_i|Y = y_i)$$

بنابراین متوسط مقدار عدم قطعیت که بعد از مشاهده Y برای X باقی می‌ماند برابر است با

$$H(X|Y) = -\sum_j P_Y(y_i) H(X|Y = y_i) \tag{۱.۱}$$

به طوری که

$$P_Y(y_i) = P(Y = y_i)$$

بنابراین آنتروپی شرطی X به شرط Y برابر است با

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

تعاریف فوق برای بیش از دو متغیر تصادفی به صورت زیر می‌باشد.

$$H(X, Y \mid Z) = -\sum_i \sum_j \sum_k P(X = x_i, Y = y_j, Z = z_k) \log P(X = x_i, Y = y_j \mid Z_k)$$

در قضایای زیر ارتباط میان آنتروپی توأم و آنتروپی حاشیه‌ای متغیرهای تصادفی را بررسی می‌کنیم.

قضیه ۸.۱ برای دو متغیر تصادفی X و Y
الف) $H(X \mid Y) \leq H(X)$ و تساوی وقتی برقرار است که X و Y مستقل باشند.
ب) $H(X,Y) \leq H(X) + H(Y)$ و تساوی وقتی برقرار است که X و Y مستقل باشند.
اثبات در مرجع [۱۷] آمده است.

نتیجه ۹.۱ برای هر n متغیر تصادفی $X_۱, X_۲, \ldots, X_n$ داریم

$$H(X_۱, X_۲, \ldots, X_n) \leq H(X_۱) + H(X_۲) + \ldots + H(X_n)$$

و تساوی برقرار است اگر و تنها اگر $X_۱, X_۲, \ldots, X_n$ مستقل باشند.
اثبات در مرجع [۱۷] آمده است.

کاربرد تعاریف و قضایای فوق را در مثال زیر ملاحظه کنید.

مثال ۱۰.۱ تاس سالمی را ۲ بار پرتاب می کنیم، از برآمدهای آن هیچ اطلاعی نداریم. مقایسه پاسخ سوالات زیر جالب خواهد بود .
۱) اندازه عدم قطعیت برآمدهای این تاس
۲) اندازه عدم قطعیت برآمدهای تاس اول و دوم
۳) اگر بدانیم برآمدهای تاس اول زوج است ، اندازه عدم قطعیت چقدر کاهش می‌یابد.
۴) اگر بدانیم یکی از برآمدها زوج است ، اندازه عدم قطعیت چقدر کاهش می‌یابد.
فرض کنید X و Y به ترتیب نشان دهنده‌ی برآمدهای تاس اول و دوم باشد. در این صورت فضای نمونه آزمایش به صورت زیر است

$$S_{XY} = \{(۱,۱), \ldots, (۱,٦), (۲,۱), \ldots, (۲,٦), \ldots, (٦,۱), \ldots, (٦,٦)\}$$

با فرض هم شانس بودن برآمدها ، جدول توزیع احتمال توأم X و Y به صورت زیر است.

$X \setminus Y$	۱	۲	۳	۴	۵	٦	$f_X(x)$
۱	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{٦}$
۲	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{٦}$
۳	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{٦}$
۴	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{٦}$
۵	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{٦}$
٦	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{۳٦}$	$\frac{۱}{٦}$
$f_Y(y)$	$\frac{۱}{٦}$	$\frac{۱}{٦}$	$\frac{۱}{٦}$	$\frac{۱}{٦}$	$\frac{۱}{٦}$	$\frac{۱}{٦}$	۱

$$-\log \frac{۱}{۲} = ۱ bit \qquad , \qquad -\log \frac{۱}{۳} = ۱/۵۸۵ bit$$

حل .

(۱)

$$\begin{aligned} H(X,Y) &= -(\frac{۱}{۳۶}\log\frac{۱}{۳۶}+\frac{۱}{۳۶}\log\frac{۱}{۳۶}+\ldots+\frac{۱}{۳۶}\log\frac{۱}{۳۶}) \\ &= -۳۶.\frac{۱}{۳۶}\log\frac{۱}{۳۶} = -\log\frac{۱}{۳۶} = ۵/۱۷۰\, bit \end{aligned}$$

یعنی متوسط اطلاعات بدست آمده در پرتاب ۲ تاس برابر ۵/۱۷۰ بیت است.

(۲)

$$\begin{aligned} H(X) &= -(\frac{۱}{۶}\log\frac{۱}{۶}+\frac{۱}{۶}\log\frac{۱}{۶}+\ldots+\frac{۱}{۶}\log\frac{۱}{۶}) \\ &= -۶.\frac{۱}{۶}\log\frac{۱}{۶} = -\log\frac{۱}{۶} = ۲/۵۸۵ bit \end{aligned}$$

به طور مشابه خواهیم داشت:

$$H(Y) = ۲/۵۸۵ bit$$

یعنی متوسط اطلاعات بدست آمده در پرتاب ۲ تاس برابر ۲/۵۸۵ بیت است.

(۳)

در این حالت فضای نمونه ما به این صورت است

$$S = \{(۲,۱),(۲,۲),\ldots,(۲,۶),(۴,۱),(۴,۲),\ldots,(۴,۶),(۶,۱),(۶,۲),\ldots,(۶,۶)\}$$

$$\begin{aligned} H(Y \mid X = \text{برآمد تاس زوج باشد}) &= -(\frac{۱}{۱۸}\log\frac{۱}{۱۸}+\frac{۱}{۱۸}\log\frac{۱}{۱۸}+\ldots+\frac{۱}{۱۸}\log\frac{۱}{۱۸}) \\ &= -۱۸.\frac{۱}{۱۸}\log\frac{۱}{۱۸} = -\log\frac{۱}{۱۸} = ۴/۱۷۰\, bit \end{aligned}$$

$$H(X,Y) - H(Y \mid X = \text{برآمد تاس اول زوج}) = ۵/۱۷۰ - ۴/۱۷۰ = ۱ bit$$

بنابراین با دانستن این که تاس اول زوج است به اندازه ۱ بیت از عدم قطعیت کاسته می‌شود.

(۴)

اگر پیشامد زوج بودن یکی از برآمدها را با A نشان دهیم، در این صورت

$$\begin{aligned} A = \{ & (۱,۲),(۱,۴),(۱,۶),(۲,۱),(۲,۲),(۲,۳),(۲,۴),(۲,۵),(۲,۶),(۳,۲),(۳,۴), \\ & (۳,۶),(۴,۱),(۴,۲),\ldots,(۴,۶),(۵,۲),(۵,۴),(۵,۶),(۶,۱),(۶,۲),\ldots,(۶,۶)\} \end{aligned}$$

و در این حالت جدول توزیع توأم به صورت زیر است

$X \setminus Y$	۱	۲	۳	۴	۵	۶	$f(X \mid A)$
۱	۰	$\frac{۱}{۲۷}$	۰	$\frac{۱}{۲۷}$	۰	$\frac{۱}{۲۷}$	$\frac{۳}{۲۷}$
۲	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۶}{۲۷}$
۳	۰	$\frac{۱}{۲۷}$	۰	$\frac{۱}{۲۷}$	۰	$\frac{۱}{۲۷}$	$\frac{۳}{۲۷}$
۴	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۶}{۲۷}$
۵	۰	$\frac{۱}{۲۷}$	۰	$\frac{۱}{۲۷}$	۰	$\frac{۱}{۲۷}$	$\frac{۳}{۲۷}$
۶	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۱}{۲۷}$	$\frac{۶}{۲۷}$
$f(Y \mid A)$	$\frac{۳}{۲۷}$	$\frac{۶}{۲۷}$	$\frac{۳}{۲۷}$	$\frac{۶}{۲۷}$	$\frac{۳}{۲۷}$	$\frac{۶}{۲۷}$	۱

$$\begin{aligned} H(X,Y \mid A) &= -(۰ + \frac{۱}{۲۷}\log\frac{۱}{۲۷} + ۰ + \frac{۱}{۲۷}\log\frac{۱}{۲۷} + \ldots + \frac{۱}{۲۷}\log\frac{۱}{۲۷}) \\ &= -\log\frac{۱}{۲۷} = ۴/۷۵۵ bit \end{aligned}$$

$$H(X,Y) - H(X,Y \mid A) = (۵/۱۷۰ - ۴/۷۵۵) bit = ۰/۴۱۵ bit$$

بنابراین با دانستن این که برآمد یکی از تاس ها زوج باشد، به اندازه ۰/۴۱۵ بیت از عدم قطعیت کاسته می‌شود.
اکنون از مثال فوق نتایج زیر بدست می آید.
الف) به دلیل استقلال X و Y

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

ب) با مقایسه جواب های (۲) و (۳) مشاهده می شود که اندازه کاهش در عدم قطعیت در برآمدهای تاس ها با اطلاع کسب شده توسط قسمت (۳)، یک بیت، و با اطلاع کسب شده توسط قسمت (۴)، ۰/۴۱۵ بیت می باشد این نتیجه دور از انتظار نیست، زیرا در قسمت (۳)، اطلاع بیشتری از برآمد تاس در دو بار پرتاب می‌دهد.

۱—۳—۳ قاعده زنجیری آنتروپی برای دو متغیر

قضیه ۱۱.۱ اگر X و Y متغیرهای تصادفی گسسته با تابع جرم احتمال توأم $P(X = x_i, Y = y_j)$ باشد، آنگاه

الف) $H(X,Y) = H(X) + H(Y \mid X)$

ب) $H(X,Y) = H(Y) + H(X \mid Y)$

اثبات در مرجع [۱۷] آمده است.

قضیه ۱۲.۱ اگر X، Y و Z متغیرهای تصادفی گسسته باشند و $H(X,Y|Z)$ آنتروپی توأم X و Y به شرط Z باشد. داریم

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

اثبات در مرجع [۱۷] آمده است.

۱—۴ آنتروپی متغیرهای تصادفی پیوسته

تعریف ۱۳.۱ فرض کنید X متغیر تصادفی پیوسته با تابع چگالی احتمال $f(x)$ باشد، آنگاه آنتروپی X به صورت زیر است

$$H(X) = -E(\log f(X)) = -\int_{-\infty}^{\infty} f(x)\log f(x)d(x)$$

به شرط آنکه انتگرال فوق موجود باشد.
برای یک توزیع با پراکنش کمتر و قله نوک تیزتر، مد یک نماد خوبی برای نشان دادن این است که مشاهدات نمونه کجا قرار می گیرند. بهر حال برای توزیع های با تیزی کمتر و پراکنش بیشتر ناحیه ای که مشاهدات نمونه ممکن است قرار گیرند می تواند کاملاً بزرگ باشد.
به طور کلی بین یک خانواده از توزیع ها ، توزیع هایی که پراکنش بیشتری دارند آنتروپی بالاتری نیز دارند.
بنابراین در بین همه توزیع هایی که دامنه آنها در فاصله $[a,b]$ قرار دارند انتظار داریم توزیع یکنواخت در این فاصله بیشترین آنتروپی را داشته باشد چون تابع صعودی از طول فاصله است.

مثال ۱۴.۱ اگر $X \sim U(a,b)$. آنتروپی X به صورت زیر محاسبه می‌شود

$$\begin{aligned} f(x) &= \frac{۱}{b-a} \\ H(X) &= E(-\log f(x)) = E(-\log \frac{۱}{b-a}) = \log(b-a) \end{aligned}$$

۱—۴—۱ آنتروپی متغیرهای تصادفی توأم

آنتروپی توأم دو متغیر تصادفی پیوسته X و Y به صورت زیر تعریف می‌شود

$$H(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log f(x,y)dxdy$$

به شرط آنکه انتگرال فوق موجود باشد

نتیجه ۱۵.۱ برای n متغیر تصادفی پیوسته $X_۱, X_۲, \ldots, X_n$ آنتروپی توأم به صورت زیر تعریف می‌شود

$$H(X_۱, X_۲, \ldots, X_n) = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_۱, \ldots, x_n) \log f(x_۱, \ldots, x_n) dx_۱ \ldots dx_n$$

به شرط آنکه انتگرال فوق موجود باشد.

۱—۴—۲ آنتروپی شرطی

آنتروپی شرطی متغیر تصادفی پیوسته X به شرط Y به صورت زیر است

$$H(X|Y) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \log f(x|y) dx dy$$

به شرط آنکه انتگرال فوق موجود باشد.
قضیه ۱۶.۱ برای دو متغیر تصادفی پیوسته X و Y که $H(X)$ و $H(Y)$ محدود باشند، داریم
الف) $H(X \mid Y) \leq H(X)$ و تساوی برقرار است اگر و تنها اگر X و Y مستقل باشند.
ب) $H(X, Y) \leq H(X) + H(Y)$ و تساوی برقرار است اگر و تنها اگر X و Y مستقل باشند.
ج) $H(X, Y) = H(X) + H(Y \mid X)$
د) $H(X, Y) = H(Y) + H(X \mid Y)$
اثبات در مرجع [۱۷] آمده است.

فصل ۲

همگرایی

۲ـ۱ مقدّمه

در این فصل در مورد همگرایی متغیّرهای تصادفی صحبت می‌کنیم. اگر $< X_n >$ دنباله‌ای از متغیّرهای تصادفی باشد در چه حالتی می‌توان گفت دنباله X_n همگرا است. در بخش دوّم در مورد همگرایی تقریباً مطمئن ، همگرایی در احتمال، همگرایی در توزیع و قضیه حدّ مرکزی، در بخش سوّم در مورد سازگاری و روش دلتا و برآوردگرهای جایگذاری (PIE) صحبت می‌کنیم.

۲ـ۲ انواع همگرایی‌ها

۲ـ۲ـ۱ همگرایی تقریباً مطمئن

در احتمال، پیشامدهایی که احتمال آنها صفر است اگر چه در بعضی مواقع ممکن است ایجاد دردسر کنند، ولی در بیشتر مواقع نیز می‌توان از آنها صرف نظر کرد. اگر مجموعه نقاطی که دنباله $< X_n >$ در آن همگرا نیست دارای احتمال صفر باشد می‌گوییم $< X_n >$ به X تقریباً مطمئن ($a.s$ [۱]) همگراست. همگرایی تقریباً مطمئن را با نماد $X_n \xrightarrow{a.s} X$ نشان می‌دهیم.
اگر $X_n \xrightarrow{a.s} X$ آنگاه $P(\omega : X_n(\omega) \to X(\omega)) = ۱$ از این رو همگرایی $a.s.$ را همگرایی با احتمال یک نیز می‌گویند.

[۱] almost sure