آنتروپی

آنتروپی

آنتروپی متغیّر تصادفی $\boldsymbol{X}$ را، که با $H(\boldsymbol{X})$ نشان داده می‌شود، کلود ای. شانون [۱۳] در مقاله‌ای مهّم از نظر تاریخی، به عنوان اندازهٔ انتخاب، عدم حتمیت، و اطلاع * معرفی کرد. فرض کنید

$$H(\boldsymbol{X}) = -E[\log P_{X}(x)],$$

که در آن، $\boldsymbol{X}$ متغیّر تصادفی گسسته‌ای است و $p_{X}(x) = P_r[\boldsymbol{X} = x]$.

اگر $p_i = P_r[\boldsymbol{X} = x_i] \geq ۰, i = ۱,۲,\ldots,n$ و $\sum_{i=۱}^{n} p_i = ۱$، در این صورت

$$\begin{aligned} H(\boldsymbol{X}) &= H(p_۱, p_۲, \ldots, p_n) \\ &= -\sum_{i=۱}^{n} p_i \log p_i, \end{aligned}$$

($۰ \log ۰$ به صورت $\lim_{p \to ۰} p \log p = ۰$ تعریف می‌شود) وقتی که بتواند تنها تعدادی متناهی از مقادیر را با احتمال مثبت بپذیرد، نمایش مرسوم برای آنتروپی $\boldsymbol{X}$ است. این تعریف به آسانی به هر تعداد متناهی از متغیّرهای تصادفی گسسته که توزیع توأم آنها، روی مجموعه‌ای متناهی از نقاط، دارای احتمال مثبت است، تعمیم داده می‌شود. به خصوص، فرض کنید $(\boldsymbol{X},\boldsymbol{Y})$ یک جفت متغیّر تصادفی با

$$\begin{aligned} P\{\boldsymbol{X} = x_i, \boldsymbol{Y} = y_j\} = p_{ij} > ۰ \quad & i = ۱,۲,\ldots,m, \\ & j = ۱,۲,\ldots,n; \end{aligned}$$

$\sum_{i=۱}^{m} \sum_{j=۱}^{n} p_{ij=۱}$ باشد. در این صورت،

$$\begin{aligned} H(\boldsymbol{X},\boldsymbol{Y}) &= H(p_{۱۱}, p_{۱۲}, \ldots, p_{mn}) \\ &= -\sum_{i=۱}^{m} \sum_{j=۱}^{n} p_{ij} \log p_{ij}. \end{aligned}$$

همچنین، آنتروپی شرطی، $H(\mathbf{X}|\mathbf{Y})$، به صورت $\sum_{j=۱}^{n} p_{.j} H(\boldsymbol{X}|\boldsymbol{Y} = y_j)$ تعریف می‌شود، که در آن

$$\begin{aligned} H(\boldsymbol{X}|\boldsymbol{Y} = y_j) &= -\sum_{i=۱}^{m} P[\boldsymbol{X} = x_i|\boldsymbol{Y} = y_j] \\ & \quad \times \log P[\boldsymbol{X} = x_i|\boldsymbol{Y} = y_j] \end{aligned}$$

و

$$p_{.j} = \sum_{i=۱}^{m} p_{ij}, \quad p_{i.} = \sum_{j=۱}^{n} p_{ij}.$$

ویژگیهای آنتروپی و یکتایی

برای توجیه استفاده از $H(p_۱, p_۲, \ldots, p_n)$ به عنوان اندازهٔ عدم حتمیت یا اطلاع، معرفی «قضیهٔ یکتایی» شانون [۱۳] کاری مطلوب است.

قضیهٔ یکتایی. فرض کنید $\boldsymbol{X}$ و $\boldsymbol{Y}$ متغیّرهای تصادفی گسستهٔ دلخواهی باشند که توزیع توأم آنها $P\{\boldsymbol{X} = x_i, \boldsymbol{Y} = y_j\} = p_{ij}, i = ۱,۲,\ldots,m,$ $j = ۱,۲,\ldots,n$ است و توزیعهای حاشیه‌ای به ترتیب با $p_{i.}$ و $p_{.j}$ مشخص شوند. برای هر عدد صحیح مثبت n و هر بردار $(p_۱, p_۲, \ldots, p_n)$ با شرایط $p_i \geq ۰$ و $\sum_{i=۱}^{n} p_i = ۱$، فرض کنید $f(p_۱, p_۲, \ldots, p_n)$ تابعی با مقادیر حقیقی باشد. اگر

الف. $f(p_۱, p_۲, \ldots, p_n) \leq f(۱/n, ۱/n, \ldots, ۱/n),$

$n = ۱,۲,\ldots,$

ب. $f(p_{11},\ldots,p_{mn}) = f(p_{.1},p_{.2},\ldots,p_{.n}) + \sum_{j=1}^{n} p_{.j} \sum_{j=1}^{m} f((p_{ij}/p_{.j}),\ldots,(p_{mj}/p_{.j}))$,

ج. $f(p_1,p_2,\ldots,p_n,0) = f(p_1,p_2,\ldots,p_n)$,

د. برای هر n، $f(p_1,p_2,\ldots,p_n)$ تابعی پیوسته از p_1،p_2،...،p_n باشد،

در این صورت برای ثابت مثبت c،

$$f(p_1,p_2,\ldots,p_n) = cH(p_1,\ldots,p_n)$$

و

$$\sum_{j=1}^{m} p_{.j} \sum_{i=1}^{m} f(\frac{p_{ij}}{p_{.j}},\ldots,\frac{p_{mj}}{p_{.j}}) = cH(\boldsymbol{X}|\boldsymbol{Y}),$$

یا $f(p_1,p_2,\ldots,p_n)$ متحد با صفر است.

پس هر چنین تابعی صرف نظر از ضریبی ثابت، با $H(p_1,p_2,\ldots,p_n)$ یکی است. امّا توجه کنید که پایهٔ لگاریتمها را مشخص نکرده‌ایم. به صورت قراردادی، تعریف می‌کنیم

$$H(\frac{1}{2},\frac{1}{2}) = -(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}) = 1,$$

که هم‌ارز با استفاده از لگاریتم به پایهٔ ۲ است. این قرارداد معادل با انتخاب یک مقیاس اندازه‌گیری برای آنتروپی است و این واحدها را بیتها(ی اطلاع) می‌نامند؛ «بیت»[1] نماد اختصاری «binary digit» [به معنای رقم دو دویی] است.

حال $H(p_1,p_2,\ldots,p_n) = 0$ اگر و تنها اگر به ازای i یی، $p_i = 1$ و $H(1/n, 1/n\ldots, 1/n) = \log n$. اگر به ازای i یی، $p_i = 1$، عدم حتمیتی وجود ندارد، و اگر هر $p_i = 1/n$، با ناحتمی‌ترین وضعیت (برای n ثابت) مواجهیم بدین مفهوم که پیشگویی برآمد چنین آزمایشی سخت‌ترین وضع ممکن را دارد. ویژگیهای (الف) و (ج) مستلزم این حکم‌اند که عدم حتمیت، تابعی یکنواصعودی از n است، که n تعداد شقهای ممکن است وقتی که این شِقها همشانس‌اند. اگر $\boldsymbol{X}_1$،$\boldsymbol{X}_2$،...،$\boldsymbol{X}_t$ متغیّرهای تصادفی مستقل باشند بنا بر ویژگی (ب)، $H(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_t) = \sum_{i=1}^{t} H(\boldsymbol{X}_i)$. به علاوه اگر، $\boldsymbol{X}_i$ ها همتوزیع باشند، $H(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_t) = tH(\boldsymbol{X}_1)$.

برای شناساندن آنتروپی به عنوان یک اندازهٔ اطلاع، آزمایشی را تجسم کنید که در آن، مشاهده‌گری متغیّر تصادفی $\boldsymbol{X}$ را مشاهده و سپس مقدار مشاهده شده را به شما منتقل می‌کند. قبل از اینکه شما این اطلاع را به دست آورید، عدم حتمیت برابر با $H(\boldsymbol{X})$ است. بعد از دریافت اطلاع، عدم حتمیت برابر با صفر است. با این تعبیر که «اطلاع» همان رفع عدم حتمیت است، $H(\boldsymbol{X})$ یک اندازهٔ طبیعی برای اطلاع است.

با استفاده از بیتها به عنوان واحد آنتروپی، $H(\boldsymbol{X})$ کران پایینی برای متوسط تعداد ارقام دو دویی (یا پاسخهای به سؤالهای آری ـ نه) مورد نیاز برای تعیین مقدار $\boldsymbol{X}$ است. برای روشن شدن این مطلب، از چند مثال زیر استفاده می‌شود.

مثال ۱. فرض کنید $\boldsymbol{X}_1$،$\boldsymbol{X}_2$،...،$\boldsymbol{X}_t$ متغیّرهای تصادفی مستقل برنولی با $P\{\boldsymbol{X}_i = 1\} = p_i$ و $P\{\boldsymbol{X}_i = 0\} = q_i$، $i = 1,2,\ldots,t$ باشند. در این صورت

$$H(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_t) = -\sum_{i=1}^{t}(p_i \log p_i + q_i \log q_i)$$

و اگر هر $p_i = \frac{1}{2}$، $H(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots,\boldsymbol{X}_t) = t$. سؤالهای آری ـ نه در این حالت عبارت‌اند از: آیا مشاهدهٔ i ام برابر «۱» است؟

مثال ۲. دوازده سکه و یک ترازوی دو کفّه تهیه شده است. می‌دانیم که یکی از سکّه‌ها تقلبی است. احتمال اینکه سکّه‌ای معین تقلبی باشد $\frac{1}{12}$ است، بنابر این عدم حتمیت عبارت است از $H(\frac{1}{12},\frac{1}{12},\ldots,\frac{1}{12}) = -12(\frac{1}{12}\log\frac{1}{12}) \sim 3/585$. هر وزن کردنی با استفاده از ترازو، سه برآمد ممکن دارد: طرف چپ سبکتر، طرف راست سبکتر، هر دو طرف برابر، پس ممکن است که برای هر وزن کردن، $\log 3$ واحد اطلاع یا ۱/۵۸۵ بیت اطلاع مشاهده کنیم، به شرطی که بتوانیم سکّه‌ها را برای وزن کردن به قسمی انتخاب کنیم که این سه حالت همشانس باشند از این مطلب، دریافت می‌شود که به ۲/۲~۳/۵۸۵/۱/۵۸۵ امتحان نیاز است. چون ۲/۲ عددی صحیح نیست، ۳ امتحان لازم است. لذا به صورت زیر عمل می‌کنیم: ۴ سکه که به تصادف انتخاب شده‌اند در کفهٔ چپ و ۴ سکه در طرف راست قرار می‌دهیم. اولین وزن کردن، مستلزم ۱/۵۸۵ بیت اطلاع است و سکه تقلبی را به صورت عضوی از یک مجموعه از ۴ سکه مشخص می‌کند. این چهار سکه را به صورت زیر تقسیم می‌کنیم: ۱ سکه در کفهٔ چپ، یکی در کفهٔ راست و ۲ تا بدون وزن کردن، لذا این عمل $-2(\frac{1}{4}\log\frac{1}{4}) - \frac{1}{2}\log\frac{1}{2} = 1/5$ بیت اطلاع و یا بعد از دوبار وزن کردن، در کل ۳/۰۸۵ بیت به دست می‌دهد.

[1] bit

با احتمال $\frac{1}{2}$، به وزن کردن دیگری نیاز نیست. در غیر این صورت از طریق آخرین وزن کردن دو سکّه باقی مانده ۰/۵ بیت اطلاع قابل حصول است. توجه کنید که متوسط تعداد وزن کردنها با استفاده از این طرح برابر است با $3>2.5=(0.5)3+(0.5)2$. توجه کنید که اطلاع کل مورد نیاز نیز $3.585=0.500+1.500+1.585$ است که مقدار آغازین عدم حتمیت است. ن. ک. **توزین، طرحهای**.

یک ویژگی مهم آنتروپی، با قضیه رمزگذاری (بدون نوفه) معلوم می‌شود. فرض کنید $\boldsymbol{X}$، متغیّری تصادفی با $P\{\boldsymbol{X}=x_i\}=p_i\geq 0$، به ازای $i=1,2,\ldots,n$، $\sum_{i=1}^{n} p_i=1$ باشد. گردایه‌ای از دنباله‌های صفرها و یکها را یک رمز دو دویی برای $\boldsymbol{X}$ می‌گویند اگر هر x_i دارای دنبالهٔ متناظری باشد که قسمت آغازین دنباله متناظر برای x_j ی با $j\neq i$، $1\leq i\leq j\leq n$، نباشد. این شرط آخر، برای انجام رمزگشایی، شرطی اساسی است. مثلاً، برای $n=3$، مجموعهٔ دنباله‌های $(11, 10, 0)$ اجازه رمزگشایی را می‌دهد امّا $(00, 1, 0)$ این اجازه را نمی‌دهد، زیرا اگر ۰۰ مشاهده شود واضح نیست که آیا این رمز به نشانهٔ x_1 است که پس از آن x_1 آمده یا به نشانهٔ x_3 است. به خصوص، رمز دو دویی برای $\boldsymbol{X}$ مربوط به امید طول مینیمال $E[L]$ در $E[L]\geq H(\boldsymbol{X})$ صدق می‌کند، یا آنتروپی، یک کران پایینی برای امید طول چنین رمزهایی است.

مثال ۳. فرض کنید $\boldsymbol{X}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)$ که در آن، $\boldsymbol{X}$ ها متغیّرهای تصادفی برنولی* مستقل و همتوزیع با $P[X_i=1]=p=1-P[X_i=0]$ هستند. اگر $p=\frac{1}{2}$، در این صورت $H(\boldsymbol{X})=3$ و دنبالهٔ برآمدها، رمز بهینه است که کران پایین را به دست می‌دهد. اگر $p=\frac{3}{4}$، $H(\boldsymbol{X})=2.434$. یک رمز پیشنهادی در جدول ۱ داده شده است.

رمز مذکور دارای $E[L]=2.649$ است، که از سه رقمی که برای توصیف برآمدهای ممکن به کار رفت کمتر است.

جدول ۱

کد	احتمال	برآمدها
۱۱۱۱۱۱۱	۰/۰۱۶	۰۰۰
۱۱۱۱۱۱۰	۰/۰۴۷	۰۰۱
۱۱۱۱۱۰	۰/۰۴۷	۰۱۰
۱۱۱۱۰	۰/۰۴۷	۱۰۰
۱۱۱۰	۰/۱۴۱	۰۱۱
۱۱۰	۰/۱۴۱	۱۰۱
۱۰	۰/۱۴۱	۱۱۰
۰	۰/۴۲۲	۱۱۱

تعریف آنتروپی بی درنگ به متغیّرهای تصادفی گسستهٔ دلخواه تعمیم می‌یابد. امّا، ساختن متغیّرهای گسستهٔ $\boldsymbol{X}$ با $H(\boldsymbol{X})=\infty$ آسان است.

آنتروپی به عنوان پارامتر آمار توصیفی

آنتروپی $\boldsymbol{X}$، یا هم‌ارز آن آنتروپی تابع جرم احتمال $p_X(x)$ را می‌توان به صورت یک کمیت توصیفی در نظر گرفت درست همان طور که میانه*، مد*، واریانس*، و ضریب چولگی* را می‌توان به صورت پارامترهای توصیفی تلقی کرد.

آنتروپی، اندازه‌ای از میزان تمرکز احتمال در چند نقطه یا پراکنده شدن آن روی نقاط بسیار است. پس آنتروپی، اندازه‌ای از پراکندگی، چیزی نظیر انحراف استاندارد است. با این حال تفاوتهایی اساسی بین آنها موجود است از این لحاظ که آنتروپی به ترتیب برآمدها وابستگی ندارد. پس، همان طور که نشان داده شده است، سه متغیّر تصادفی که در جدول ۲ آورده‌ایم آنتروپیهای همانند ولی واریانسهای اساساً متفاوت، دارند.

به طور مشخص، آنتروپی تحت هر تبدیل یک به یک مقادیر x_i، $i=1,2,\ldots,n$ ناورداست. این ویژگی، آنتروپی را به خصوص اندازهٔ توصیفی مفیدی برای متغیّرهای تصادفی می‌سازد که مقادیر غیر عددی می‌پذیرند. بدین لحاظ، آنتروپی غالباً برای داده‌های مربوط به زبان، نظیر فراوانیهای واژه‌ها در یک زبان به کار می رود. بر حسب مدلهای احتمالاتی، آنتروپی به عنوان اندازهٔ ناهمگنی برای توزیعهای چندجمله‌ای* (با خانه‌های نامرتب) بهترین تعبیر را دارد.

جدول ۲

	متغیر		
x	$\mathbf{P}_{X_1}(x)$	$\mathbf{P}_{X_2}(x)$	$\mathbf{P}_{X_3}(x)$
-2	0	$\frac{1}{4}$	0
-1	$\frac{1}{4}$	0	0
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
1	$\frac{1}{4}$	0	$\frac{1}{4}$
2	0	$\frac{1}{4}$	$\frac{1}{2}$

$$H(\boldsymbol{X}_1)=H(\boldsymbol{X}_2)=H(\boldsymbol{X}_3)=1.5$$
$$E(\boldsymbol{X}_1)=E(\boldsymbol{X}_2)=0, E(\boldsymbol{X}_3)=1.25$$
$$\sigma^2_{X_1}=0.5, \sigma^2_{X_2}=2, \sigma^2_{X_3}=0.6875.$$

سودمند بودن آنتروپی به عنوان اندازه‌ای توصیفی، توجه به برآورد آماری آن را از روی داده‌های مشاهده شده، توجیه می‌کند. بازارین [۳] برآوردگر ماکسیمم درستنمایی

را مطالعه کرده است. بخش مفصلتری از مسئله برآورد در هاریس [۵] داده شده است.

آنتروپی متغیّرهای تصادفی پیوسته

برای متغیّرهای تصادفی پیوسته، آنتروپی متغیّر تصادفی $\boldsymbol{X}$ به قیاس، به صورت $H(\boldsymbol{X}) = -E[\log f_X(\mathbf{X})]$ تعریف می‌شود، که در آن $f_X(x)$ معّرف تابع چگالی احتمالی متغیّر تصادفی $\boldsymbol{X}$ است. در اینجا استفاده از یک مقیاس دیگر اندازه‌گیری و استفاده از لگاریتمهای طبیعی به جای لگاریتمهای به پایهٔ ۲، راحت‌تر است.

به عنوان مثال، برای توزیع نمایی* با

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0, \\ H(\boldsymbol{X}) &= \log(e/\lambda) = 1 - \log\lambda. \end{aligned}$$

همچنین، برای توزیع نرمال با میانگین صفر، یعنی

$$f_X(x) = (2\pi)^{-1/2}\sigma^{-1}e^{-x^2/2\sigma^2}, \qquad \sigma > 0,$$

داریم

$$\begin{aligned} H(\boldsymbol{X}) &= \log[(2\pi e)^{1/2}\sigma] \\ &= \log[(2\pi)^{1/2}\sigma] + \frac{1}{2}. \end{aligned}$$

مثالهای بالا این را که دیگر ویژگی ناوردایی برای متغیّرهای تصادفی گسسته قابل اعمال نیستند، آشکار می‌کنند. امّا، بسیاری از ویژگیهای دیگر هنوز صادق‌اند یا مشابه‌هایی طبیعی دارند. برخی از اینها در زیر آمده‌اند.

۱. فرض کنید $f_X(x)$ هر تابع چگالی احتمال* با $P[a \le \boldsymbol{X} \le b] = 1$ باشد. در این صورت $H(\boldsymbol{X})$ ماکسیمم است هرگاه $f_X(x)$ توزیع یکنواخت* روی $[a,b]$ باشد.

۲. اگر $\boldsymbol{X}$ و $\boldsymbol{Y}$ متغیّرهایی تصادفی با تابع چگالی احتمال توأم پیوستهٔ $f_{X,Y}(x,y)$ باشد، آن‌گاه

$$H(\boldsymbol{X},\boldsymbol{Y}) \le H(\boldsymbol{X}) + H(\boldsymbol{Y})$$

برقرار است، اگر و تنها اگر $\boldsymbol{X}$ و $\boldsymbol{Y}$ مستقل باشند.

۳. اگر $\boldsymbol{X}$ متغیّری تصادفی با $P\{\boldsymbol{X} > 0\} = 1$ و $E[\boldsymbol{X}] = \alpha$ باشد، آن‌گاه آنتروپی ماکسیمم است هرگاه

$$f_X(x) = \alpha^{-1}e^{-x/\alpha}, \quad \alpha > 0, \quad x > 0.$$

۴. اگر $\boldsymbol{X}$ متغیّر تصادفی با $E[\boldsymbol{X}] = 0$ و واریانس مفروض σ^2 باشد، آن‌گاه آنتروپی ماکسیمم است اگر و تنها اگر $\boldsymbol{X}$ متغیّری تصادفی با توزیع نرمال باشد. این ویژگی آخر در ساختن یک آزمون نرمال بودن* به وسیله واسیچک [۱۵] به کار رفته است.

جنبه‌های گوناگون

در استنباط آماری، چند اندازه را در رابطهٔ نزدیک با آنتروپی مفید یافته‌اند. بین پرکاربردترین چنین اندازه‌هایی می‌توان از اعداد اطلاع کولبک ـ لیبلر $I(1,2)$ و $J(1,2)$ نام برد که به صورت

$$I(1,2) = \int \log[f_{X_1}(x)/f_{X_2}(x)]f_{X_1}(x)dx$$

و

$$J(1,2) = I(1,2) + I(2,1)$$

تعریف شده‌اند. با تعریف $-E[\log\ g(\boldsymbol{X})]$، که در آن $g(x)$ تابع چگالی احتمال* است، و $\boldsymbol{X}$ توسط $f_X(x)$ به عنوان «عدم دقت» آن توزیع شده است، از نتایج معروف این است که

$$-E[\log\ g(X)] \ge -E[\log f(X)],$$

برابری برقرار است، اگر و تنها اگر $g(x) = f(x)$ باشد. بنابر این $I(1,2)$ تفاضل بین عدم دقت و آنتروپی است. اگر این تفاضل بزرگ باشد، تفاوتهای آماری اساسی بین دو جامعه‌ای که توزیع آنها با $f_{X_1}(x)$ و $f_{X_2}(x)$ نشان داده شده است وجود دارند. اگر این تفاوت کوچک باشد، بیان اینکه داده‌ها از یک جامعه یا جامعه دیگر می آید مشکل خواهد بود. پس، توان* آزمونهای آماری بستگی به بزرگی $I(1,2)$ دارد. باید توجه کرد که $I(1,2)$ مقدار امیدریاضی لگاریتم نسبت درستنمایی* است؛ $J(1,2)$ اندازهٔ متقارن شدهٔ این مفهوم فاصلهٔ آماری* بین دو جامعه است.

ری‌نی [۱۱]، مفهوم آنتروپیهای مرتبهٔ α را به صورت

$$H_\alpha(p_1,p_2,\ldots,p_n) = (1-\alpha)^{-1}\log(\sum p_i^\alpha), \alpha \ne 1$$

تعریف کرده است. وقتی $\alpha \to 1$ ، $H_\alpha(p_1,p_2,\ldots,p_n)$ به $H(p_1,p_2,\ldots,p_n)$ تبدیل می‌شود. $H_2(p_1,p_2,\ldots,p_n)$ را غالباً به عنوان اندازهٔ تنّوع به کار برده‌اند. آنتروپیهای مرتبهٔ α، در هارت [۶]، با اندازه‌های توصیفی مختلف دیگر که اندازه‌های تنوع اقتصادی‌اند، مقایسه شده‌اند. خواننده باید کمپ [۷] را نیز که، دربارهٔ مطلوبیت آنتروپی مرتبهٔ دو به عنوان اندازهٔ توصیفی است، ببیند.

در نظریهٔ ارتباطات* نیز از مفهوم آنتروپی به صورت گسترده استفاده شده است، که در آن، آنتروپی به عنوان وسیلهٔ اندازه‌گیری میزان اطلاعی به کار می‌رود که می‌توان روی کانالهای ارتباطاتی نوفه‌دار ارسال کرد.

مطالب مرجعی دیگر در مرجعهای ۱،۲،۴،۸—۱۰،۱۲، و ۱۴، همچنین در مرجعهایی که قبلاً ذکر کردیم و در مدخلهای مرتبط مختلف این فرهنگ دانشنامگی داده شده‌اند.

مرجعها

[1] Abramson, N. (1963).*Information Theory and Coding*. McGraw-Hill, New York.

(آشنایی مقدماتی.)

[2] Aczél, J. and Daróczy, Z. (1975). *On Measures of Information and Their Characterizations*. Academic Press, New York.

(رفتار خیلی تخصصی اندازه‌گیری اطلاعات و ویژگیهای اندازه‌های اطلاعات گوناگون.)

[3] Bašarin, G. P. (1959). *Teor. verojatn. ee Primen.*, **4**, 361-364.

[4] Guiasu, S. (1977). *Information Theory With Applications*. McGraw-Hill, New York.

(رفتار نظری کلی نظریه اطلاع و کاربرد آن، شامل نظریه ارتباطات، استنباط آماری، نظریه کدگذاری، و تشخیص الگو.)

[5] Harris, B. (1975). *Colloq. Math. Societatis János Bolyai*, **16**, 323-355

[6] Hart, P. E. (1975). *J. R. Statist. Soc. A*,**138**, 423-434.

(شامل فهرست وسیعی از منبعهای مقاله‌هاست که اندازه‌های انحراف توصیفی مختلفی را مورد بحث قرار می‌دهد.)

[7] Kemp, A. W. (1973). *Bull. Inst. Int. Statist.*, **45**, 45-51.

[8] Kemp, A. W. (1975). *Bull. Inst. Int. Statist.*, **46**, 446-452.

[9] Khinchin, A. I. (1957). *Mathematical Foundations of Information Theory*. Dover, New York.

(گسترش دقیق ایده‌های اصلی شانون.)

[10] Kullback, S. (1959).*Information Theory and Statistcs*. Wiley, New York.

(استفاده از اندازه‌های اطلاعات در استنباط آماری، با تأکیدی خاص بر جدولهای پیشایندی و تحلیل چندمتغیّره.)

[11] Rényi, A. (1966). *Wahrscheinlichkeitsrechnung miteinem Anhang über Informationstheorie*. VEB Deutscher Verlag der Wissenschaften, Berlin.

(رفتار احتمالاتی اندازه‌های اطلاعات و ویژگیهای آنها در یک ضمیمه کتاب درسی استاندارد در نظریهٔ احتمال.)

[12] Reza, F. M. (1961). *An Introduction to Information Theoty*. McGraw-Hill, New York.

(کتاب درسی بسیار خواندنی در نظریه اطلاع که در سطح متوسط دشواری نوشته شده است.)

[13] Shannon, C. E. (1948). *Bell Syst. Tech. J.*, **27**, 379-423, 623-656.

(مقالهٔ تاریخی که در آن مفاهیم آنتروپی، ظرفیت کانال، و کدگذاری توسط شانون گسترش داده شده است.)

[14] Shannon, C. E. and Weaver, W. (1959) *the Mathematical Theory of Communication*. University of Illinois Press, Champaign, Ill.

(چاپ مجدد مقاله توسط شانون [۱۳] با توضیحات اضافه شده توسط دبلیو. وی‌وِرْ.)

[15] Vasicek, O. (1976). *J. R. Statist. Soc. B*,**38**, 54-59.

(ارتباطات، نظریهٔ آماری
چندگونگی، شاخصهای
اطلاع و کدگذاری، نظریه‌های)

ن : بی. هاریس[۱]
م : علی عمیدی

[۱] B. Harris