情報エレクトロニクス学科共通科目・2年次・夏ターム〔必修科目〕

講義「情報理論」

第3回

第2章 情報量とエントロピー(2.4～2.7節)

[復習]確率変数のエントロピー

確率変数Xの取りうる値を$a_1, a_2, \cdots, a_M$とし、Xがそれぞれの値をとる確率が$p_1, p_2, \cdots, p_M$ （ただし、$p_1+p_2+\cdots+p_M=1$)であるとき、**確率変数XのエントロピーH(X)**は、

$$\mathrm{H}(X) = -\sum_{i=1}^{M} p_i \log_2 p_i$$

と定義される。

Xのエントロピー
= Xのあいまいさ
= Xの実現値を知ることにより減るあいまいさの量
= Xの値を知ることにより得られる平均情報量

[復習]どっちの天気情報の方が価値が高い？

[1月の札幌]

	晴	曇	雨	雪
割合(%)	5.5	1.2	0.2	93.1
自己情報量(ビット)	4.18	6.38	8.97	0.10

[6月の東京]

	晴	曇	雨	雪
割合(%)	32.1	27.6	40.3	0
自己情報量(ビット)	1.63	1.85	1.31	∞

東京の6月の天気を*X*, 札幌の1月の天気を*Y*とすれば

$$H(X) = -0.055\log_2 0.055 - 0.012\log_2 0.012 - 0.002\log_2 0.002 - 0.931\log_2 0.931$$
$$=0.055\times 4.18 + 0.012\times 6.38 + 0.002\times 8.97 + 0.931\times 0.10$$
$$=0.4175$$

$$H(Y) = -0.321\log_2 0.321 - 0.276\log_2 0.276 - 0.403\log_2 0.403$$
$$=0.321\times 1.63 + 0.276\times 1.85 + 0.403\times 1.31$$
$$=1.56176$$

東京の6月の天気の方が情報量が多い！

[余談]前回の演習問題に関する疑問

[疑問] $H(X)=-\sum_{i=1}^{M} p_i \log_2 p_i$ の最大値が $\log_2 M$ であることを示すのにシャノンの補助定理を、$q_1=q_2=\cdots=q_M=1/M$ として適用して $-\sum_{i=1}^{M} p_i \log_2 p_i \leqq \log_2 M$ を導き、補助定理の等号成立条件より $p_1=p_2=\cdots=p_M=1/M$ のとき $H(X)=\log_2 M$ の最大値をとることを示した。$q_1,q_2,\ldots,q_M$ に他の値を入れたらダメなの？

シャノンの補助定理

$p_1,p_2,\cdots,p_M$ および $q_1, q_2, \cdots, q_M$ を

$$p_1+p_2+\cdots+p_M = 1,$$
$$q_1 + q_2 + \cdots + q_M \leqq 1$$

を満たす任意の非負の数とする（ただし、$p_i\neq0$ のときは $q_i\neq0$ とする）。
このとき、

$$-\sum_{i=1}^{M} p_i \log_2 q_i \geqq -\sum_{i=1}^{M} p_i \log_2 p_i$$

が成立する。

等号は $q_i = p_i$ （$i = 1, 2, \cdots, M$）のとき、またそのときに限って成立する。

[余談]答

他の値を入れるといずれかのp_iが消えず、定数で上から抑えることができない。$q_1=q_2=\cdots=q_M=1/M$とすると任意の$p_1,\ldots,p_M$に対し

$$-\sum_{i=1}^{M} p_i \log_2 p_i \leqq -\sum_{i=1}^{M} p_i \log_2 (1/M) = -\log_2 (1/M) \underbrace{\sum_{i=1}^{M} p_i}_{=1}$$

で、うまくすべてのp_iが消えている。

例えば、$M=2$、$p_1=p$, $p_2=1-p$のとき、$q_1=q$, $q_2=1-q$としてシャノンの補助定理を適用すると

$$\mathcal{H}(p) \leqq -p\log_2 q-(1-p)\log_2(1-q)=p\log_2((1-q)/q)-\log_2(1-q)$$

が任意の$0\leqq p\leqq 1$に対して成り立つ。これは、定数qに対し、エントロピー関数$\mathcal{H}(x)$は、直線$y=x\log_2((1-q)/q)-\log_2(1-q)$以下であることを示している。

[余談]答(つづき)

直線

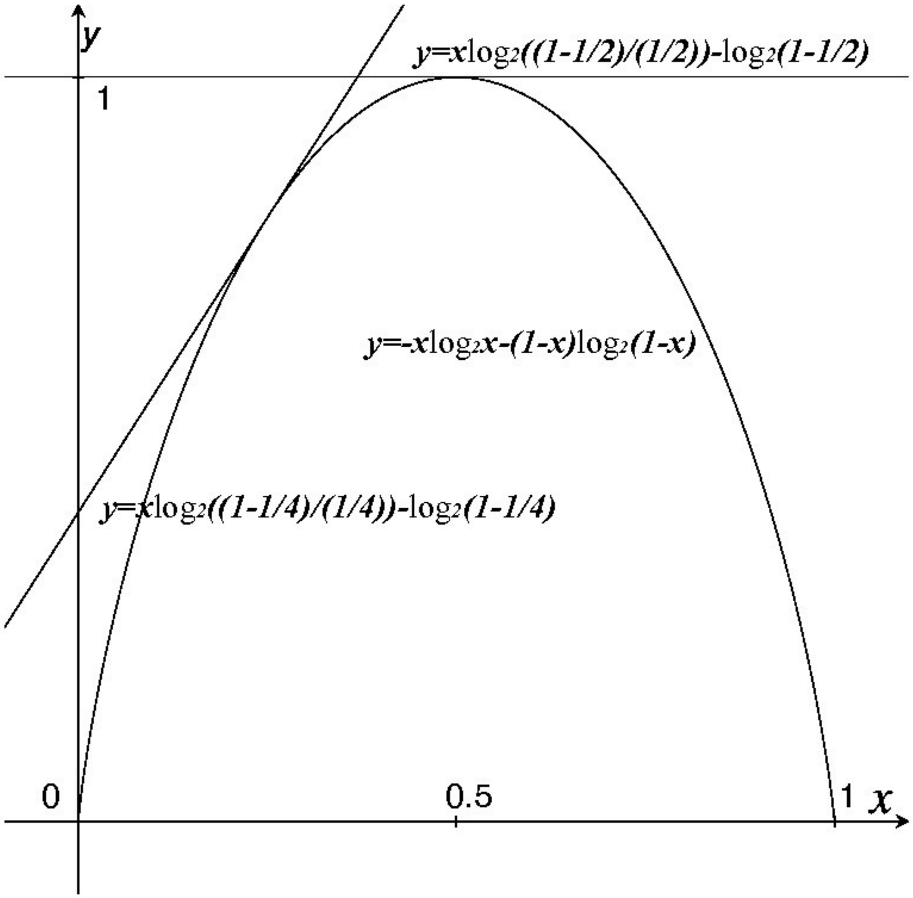
$$y = x\log_2((1-q)/q) - \log_2(1-q)$$

はx=qの座標でエントロピー関数の曲線

$$y = -x\log_2 x - (1-x)\log_2(1-x)$$

と接する直線となる。

q=1/2のとき、xに依存しない定数値をとる直線y=1になる。



結合エントロピー

確率変数XとYが独立 → XとYの値を同時に知ることによって
$H(X)+H(Y)$の平均情報量を得る

XとYが独立でない場合は？

[例]晴と雨の2値の値をとる天気Xと、
1万円以上と1万円未満の2値の値を
とるアイスクリームの売り上げYが、
右図のような分布に従っているとする。
このとき、天気Xとアイスクリームの
売上Yを同時に知ることにより得られる平均情報量は？

XとYの結合確率分布とそれぞれの確率分布

$P(x,y)$		Y: 1万円以上	Y: 1万円未満	$P(x)$
X	晴	0.5	0.1	0.6
	雨	0.2	0.2	0.4
$P(y)$		0.7	0.3	

結合エントロピー(つづき)

確率変数Zを

$Z=(X,Y)\in$ {(晴,1万円以上),(晴,1万円未満),
(雨,1万円以上),(雨,1万円未満)}

とすれば、XとYの値を同時に知ることによって得られる平均情報量は

XとYの結合確率分布とそれぞれの確率分布

$P(x,y)$		Y 1万円以上	Y 1万円未満	$P(x)$
X	晴	0.5	0.1	0.6
X	雨	0.2	0.2	0.4
$P(y)$		0.7	0.3	

$H(Z)$ = −P(晴,1万円以上)$\log_2$P(晴,1万円以上)
−P(晴,1万円未満)$\log_2$P(晴,1万円未満)
−P(雨,1万円以上)$\log_2$P(雨,1万円以上)
−P(雨,1万円未満)$\log_2$P(雨,1万円未満)

$= -0.5\log_2 0.5 - 0.1\log_2 0.1 - 0.2\log_2 0.2 - 0.2\log_2 0.2 \approx 1.76$ (bit)

XとYの結合エントロピー　$H(X,Y) = -\sum_x \sum_y P(x, y)\log_2 P(x, y)$

結合エントロピーの性質

$0 \leqq H(X,Y) \leqq H(X)+H(Y)$

$H(X,Y)=H(X)+H(Y)$はXとYが独立のときのみ成立

（証明）定義より$0 \leqq H(X,Y)$は明らか。$H(X,Y) \leqq H(X)+H(Y)$を証明する。

$$H(X)=-\sum_x p(x)\log_2 p(x)=-\sum_x \sum_y p(x,y)\log_2 p(x) \qquad (\because \sum_y p(x,y)=p(x))$$

$$H(Y)=-\sum_y p(y)\log_2 p(y)=-\sum_x \sum_y p(x,y)\log_2 p(y) \qquad (\because \sum_x p(x,y)=p(y))$$

よってシャノンの補助定理より

$$H(X)+H(Y)=-\sum_x \sum_y p(x,y)\log_2 p(x)\,p(y) \geqq -\sum_x \sum_y p(x,y)\log_2 p(x,y)=H(X,Y)$$

等号成立はすべてのx,yに対して$p(x,y)=p(x)\,p(y)$のとき、
つまりXとYが独立のときのみである。

条件付きエントロピー

- アイスクリームの売り上げYで条件付けた天気Xの確率は $P(x|y)=P(x, y)/P(y)$ より表のようになる。

Yで条件付けた
Xの条件付分布

$P(x \mid y)$		Y 1万円以上	Y 1万円未満
X	晴	5/7	1/3
	雨	2/7	2/3

- アイスクリームの売上が1万円以上であったときいた後、天気のあいまいさは

$$H(X \mid 1万円以上) = \mathcal{H}(5/7) = 0.8631 \text{ (bit)}$$

というエントロピーで表されるはずである。

$\mathcal{H}(x) = -x\log_2 x - (1-x)\log_2(1-x)$

- 同様にアイスクリームの売上が1万円未満であったときいた後、天気のあいまいさは

$$H(X \mid 1万円未満) = \mathcal{H}(1/3) = 0.9183 \text{ (bit)}$$

である。

条件付きエントロピー（つづき）

- アイスクリームの売り上げが、1万円以上、1万円未満である確率は、それぞれ 0.7、0.3であるから、あいまいさを平均すると、

Yで条件付けた
Xの条件付分布

$P(x \mid y)$		Y: 1万円以上	Y: 1万円未満
X	晴	5/7	1/3
	雨	2/7	2/3

$$H(X \mid Y) = 0.7 \times 0.8631 + 0.3 \times 0.9183 \approx 0.8797 \text{ (bit)}$$

となる。　$H(X|Y)$はYの情報を得た後のXの平均のあいまいさ

Yで条件をつけたXの条件付エントロピー $H(X \mid Y)$

$$H(X|Y) \overset{\text{def}}{=} -\sum_{y} P(y) \sum_{x} P(x|y) \log_2 P(x|y) = -\sum_{x}\sum_{y} P(x, y) \log_2 P(x|y)$$

条件付きエントロピーの性質

(1) $H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y)$
(2) $0\leqq H(X|Y)\leqq H(X)$ ($H(X|Y)=H(X)$はXとYが独立のときのみ成立)
(3) $0\leqq H(Y|X)\leqq H(Y)$ ($H(Y|X)=H(Y)$はXとYが独立のときのみ成立)

(証明)(1)
$$
\begin{aligned}
H(X,Y) &= -\sum_x \sum_y P(x,y)\log_2 P(x,y) \\
&= -\sum_x \sum_y P(x,y)\log_2 P(x)P(y|x) \\
&= -\sum_x \sum_y P(x,y)\log_2 P(x) - \sum_x \sum_y P(x,y)\log_2 P(y|x) \\
&= -\sum_x P(x)\log_2 P(x) - \sum_x P(x)\sum_y P(y|x)\log_2 P(y|x) = H(X)+H(Y|X)
\end{aligned}
$$

p(x,y)=p(x)p(y|x)

$H(X,Y)=H(Y)+H(X|Y)$も同様。

(2) $0\leqq H(X|Y)$は定義より明らか。$H(X|Y)\leqq H(X)$を証明する。

(1)より

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

$H(X,Y)\leqq H(X)+H(Y)$であるから

$$H(X|Y)\leqq H(X)$$

等号成立は$H(X,Y)=H(X)+H(Y)$のときであるから、XとYが独立のとき

(3) (2)と同様。

アイスクリームの売上から得られる情報量は？

[問題]アイスクリームの売上Yから得られる
天気Xに関する平均情報量はどのくらいか。

YのXに関する平均情報量
=**Yの値を知ることによるXに関する**
あいまいさの減少量。

XとYの結合確率分布とそれぞれの確率分布

$P(x,y)$		Y: 1万円以上	Y: 1万円未満	$P(x)$
X	晴	0.5	0.1	0.6
	雨	0.2	0.2	0.4
$P(y)$		0.7	0.3	

天気Xのエントロピー $H(X)$は

$$H(X) = \mathcal{H}(0.6) = 0.9710 \text{ (bit)}$$

Yの値を知ることによりXのエントロピーはどれだけ減少するか？

相互情報量

- アイスクリームの売上Yをきいたとき、残っている天気Xに関するあいまいさは、Yで条件をつけたXの条件付エントロピー

$$H(X \mid Y) = 0.8797 \text{ (bit)}$$

であるから、結局、アイスクリームの売上Yを知ることにより、天気Xについて

$$\begin{aligned} I(X;Y) &= H(X) - H(X \mid Y) \\ &= 0.9710 - 0.8797 = 0.0913 \text{ (bit)} \end{aligned}$$

だけ、あいまいさが減少することになる。

- すなわち、アイスクリームの売上によって$I(X;Y)=0.0913$ ビットだけの平均情報量が得られることを意味する。

XとYの相互情報量(mutual information) $I(X;Y)$

$$I(X;Y) \overset{\text{def}}{=} H(X) - H(X \mid Y)$$

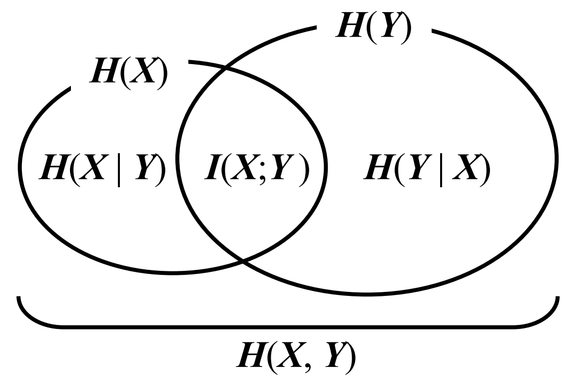
XとYの相互情報量はYの値を知ることにより得られるXに関する平均情報量

相互情報量の性質(1)

(1) $I(X;Y) = H(X) - H(X\,|\,Y) = H(Y) - H(Y\,|\,X)$

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

(2) $0 \leqq I(X;Y) \leqq \min\{H(X), H(Y)\}$



相互情報量とエントロピーの関係

[(1)の証明] 相互情報量 $I(X;Y) = H(X) - H(X\,|\,Y)$ に

$$H(X) = -\sum_x P(x)\log_2 P(x) = -\sum_x\sum_y P(x, y)\log_2 P(x)$$

および

$$H(X|Y) = -\sum_y P(y)\sum_x P(x|y)\log_2 P(x|\,y) = -\sum_x\sum_y P(x, y)\log_2 P(x|\,y)$$

を代入すると、

$$I(X;Y) = \sum_x\sum_y P(x, y)\log_2 \frac{P(x\,|\,y)}{P(x)} = \sum_x\sum_y P(x, y)\log_2 \frac{P(x, y)}{P(x)P(y)}$$

$P(x\,|\,y)p(y) = P(x,y)$

を得る。この式はXとYに関し、全くの対称である。したがって、

$$I(X;Y) = I(Y; X) = H(Y) - H(Y\,|\,X)$$

となることが分かる。さらに$H(X,Y)=H(X)+H(Y|X)$であるから以下の式も成り立つ。

$$I(X;Y) = H(X) + H(Y) - H(X, Y)$$

相互情報量の性質(2)

[(2)の証明] $0 \leqq H(X|Y) \leqq H(X)$であるから

$$0 \leqq H(X) - H(X|Y) = I(X;Y)$$

等号成立は$H(X|Y)=H(X)$のとき、つまりXとYが独立（$P(x, y) = P(x)\,P(y)$）のとき。

$H(X \mid Y), H(Y \mid X) \geqq 0$であるから

$$I(X;Y) \leqq \min\{H(X), H(Y)\}$$

相対エントロピー

同じ空間上の分布Pの分布Qに対する**相対エントロピー**※ $D(P \parallel Q)$

$$D(P \parallel Q) \stackrel{\text{def}}{=} \sum_{x} P(x) \log \frac{P(x)}{Q(x)}$$

※カルバック･ライブラー情報量とも言う

相対エントロピーは分布Pが分布Qからどれだけ遠いかを表す指標

相対エントロピーの性質

(1)$D(P \parallel Q) \geq 0$

(2) $D(P \parallel Q) = 0 \iff P = Q$

(3) $D(P \parallel Q) \neq D(Q \parallel P)$

相互情報量との関係

確率変数XとYは独立からどれだけ遠いか？

$$I(X;Y) = \sum_{x} \sum_{y} P(x,y) \log \frac{P(x,y)}{P(x)p(y)} = D(\{P(x,y)\} \parallel \{P(x)P(y)\})$$