北海道大学
Hokkaido University

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」(クラスC)

第2回

第2章 情報量とエントロピー(前半:2.1～2.3)

確率統計の基礎(独立と排反)

- 事象AとBが独立⇔$P(A \cap B)=P(A)P(B)$

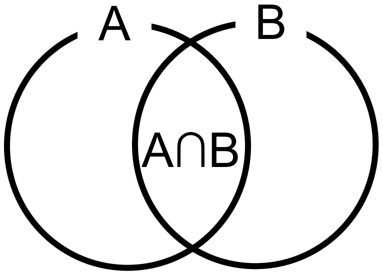
（AとBが共に起こる確率は各々が起こる確率の積になる）

- 事象AとBが排反⇔$A \cap B=\phi$⇔$P(A \cap B)=0$

（AとBは同時には起こらない）

- 確率の加法定理

$$P(A \cup B)=P(A)+P(B)-P(A \cap B)$$



事象AとBが排反ならば

$$P(A \cup B)=P(A)+P(B)$$

二項分布B(n,p)

- １回の試行において事象Aが起こる確率がpであるとき、その試行を独立にn回繰り返したとき、Aが起こる回数を確率変数XとするとXは二項分布B(n,p)に従う。このとき

$$P(X=k)={}_nC_k p^k(1-p)^{n-k}$$

- １ビットが誤る確率pの通信路にnビット送ったときのnビットの内に誤りが生じる確率は

$$P(X>0)=1-P(X=0)=1-(1-p)^n=np-{}_nC_2p^2+{}_nC_3p^3-\cdots-(-1)^np^n$$

$$\fallingdotseq np \quad \text{（pが0に近いとき）}$$

情報量はどう定義すべきか？

- よくあることが起こったという情報・・・・・価値が低い
  - 必ず(確率1で)起こることが起こったと知らされる → 得られる情報量は 0
- めったに起こらないことが起きたという情報・・・・・価値が高い
  - ほとんど起こらない(確率が 0 に近い)ことが起きたと知らされる

[1月の札幌] → 情報量は大！

Case 1

外の天気を知らない場合　情報量は小

Case 2

外の天気を知らない場合　情報量は大

Case 3

外の天気を知っている場合　情報量は0

一つの結果を知ったときに得る情報量

確率 p で起こることが起こったと知った場合に、得られる情報量を $I(p)$ とすれば、$I(p)$はどんな条件を満たすべきか？

確率が大きいほど情報量は小さい → $I(p)$ は p の単調減少関数である

２つの独立な事象の生起を同時に知ったときに得られる情報量は、片方ずつ知ったときに得られる情報量の和 → 確率p_1, p_2で起こる二つの互いに独立な事象の結合確率p_1p_2について

$$I(p_1p_2)=I(p_1)+I(p_2)$$

確率に差がなければ情報量も差がない → $I(p)$ は p の連続関数である

- これらを満たす関数 $I(p)$ は

$$I(p) = -log_a p$$

という形しかありえない（a は1より大きい定数）。

- $I(1/2)=1$ とすると $a=2$ となり、情報量 $I(p)=-\log_2 p$ となる

情報量の定義

[定義] 確率pで生起する事象が起こったことを知ったとき、得られる情報量$\boldsymbol{I}(p)$を自己情報量(self-information)とよび

$$\boldsymbol{I}(p) = -\log_a p$$

と定義する。ただしaは1より大きい定数とする。

情報量の単位は

a=2のとき　ビット(bit)

a=eのとき　ナット(nat)

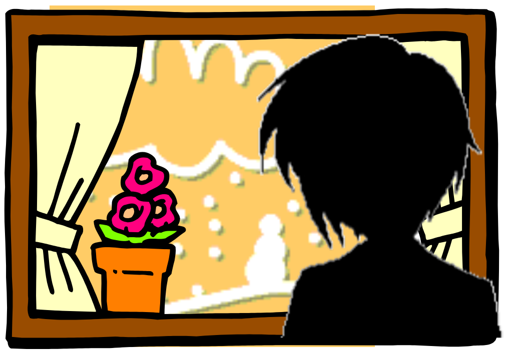
a=10のとき　デシット(decit) or ハートレー(Hartley)

1bitの情報量=確率1/2で生起する事象が起こったことを知ったとき得られる情報量

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	∞

平均情報量

M個の互いに排反な事象$a_1, a_2,\cdots, a_M$ が起こる確率が、$p_1, p_2,\cdots, p_M$ （ただし、$p_1+p_2+\cdots+p_M=1$)であるとき、M個の内の１つの事象が起こったことを知ったときに得られる**平均情報量**は

$$\bar{I} = -\sum_{i=1}^{M} p_i \log_2 p_i$$

である。

平均とは期待値のことであるから上式は

$$\bar{I} = \sum_{i=1}^{M} p_i \, I(p_i)$$

より求められている。

平均情報量の計算例

[1月の札幌の天気を知ることにより得られる平均情報量]

	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

$\bar{I} = -0.055\log_2 0.055 - 0.012\log_2 0.012 - 0.002\log_2 0.002 - 0.931\log_2 0.931$

$= 0.055\times 4.18 + 0.012\times 6.38 + 0.002\times 8.97 + 0.931\times 0.10$

$= 0.4175$(もう少し厳密に計算すると0.4207)

6月の東京の天気を知ることにより得られる平均情報量と比べてどちらが大きいか？

エントロピー（＝平均情報量）

X:１月の札幌の天気を表す確率変数、X∈{晴,曇,雨,雪}

１月の札幌の天気を知ることにより得られる平均情報量$\bar{I}$

$$= -\sum_{x\in\{\text{晴},\text{曇},\text{雨},\text{雪}\}} P\{X=x\}\log_2 P\{X=x\}$$

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。

あいまいさの尺度としてのエントロピー

- エントロピーは本来熱力学の用語であり、力学系の無秩序さを表す尺度として用いられる。
- 情報理論におけるエントロピーも、同様に無秩序さ、いいかえれば、あいまいさを表す尺度である。
  - 確率変数Xのエントロピー$H(X)$は、試行の結果（Xの実現値）を知る以前に、Xついて我々が持つ知識のあいまいさの尺度である。
  - Xの取りうる値各々の発生確率は知っているが、試行の結果がどうなるかは確定できない（あいまいさがある）。
  - 試行の結果を知った時点でその（$H(X)$の量の）あいまいさが消失する

    → あいまいさが $H(X)$ から 0 へと変化

    → 得られた情報量 $I(X) = H(X) - 0 = H(X)$

「確率変数Xの情報量」
　＝「その情報（Xの実現値）を受け取ることによるエントロピーの減少量」

エントロピーの最小値と最大値

- M個の値$a_1, a_2, \cdots, a_M$をそれぞれ$p_1, p_2, \cdots, p_M$の確率でとる確率変数Xを考える。Xのエントロピーは

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

である。$0 \leqq p_i \leqq 1$なので、明らかに

$$0 \leqq H(X)$$

であり、この式の等号が成立するのは、$p_1, p_2, \cdots, p_M$のうち一つが1で他が0の場合である。

ある一つの値しかとらないことがわかっている。

- 次のページの補助定理より

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

を得る。この式の等号が成立するのは$p_i = 1/M$のときである。

すべての値を等しい確率でとる。

M個の値をとる確率変数Xのエントロピーは次を満たす

$$0 \leqq H(X) \leqq \log_2 M$$

シャノンの補助定理

$-\sum_{i=1}^{M} p_i \log_2 p_i \leqq \log_2 M$を証明するために、次の補助定理を使う。

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、$p_i \neq 0$ のときは $q_i \neq 0$ とする）。

このとき、

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

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。

シャノンの補助定理の証明

（証明） 式①の右辺から左辺を引いた結果を$\boldsymbol{D}$とおくと

$$\boldsymbol{D} = -\sum_{i=1}^{M} p_i \log_2 p_i + \sum_{i=1}^{M} p_i \log_2 q_i$$

$$= \sum_{i=1}^{M} p_i \log_2 \frac{q_i}{p_i} = \sum_{i=1}^{M} \frac{p_i}{\ln 2} \ln \frac{q_i}{p_i}$$

$\ln x = \log_e x$

となるので $\boldsymbol{D} \leqq 0$ を示せばよい。よく知られた不等式

$$\ln x \leqq x - 1$$

を用いると、

$$\boldsymbol{D} \leqq \sum_{i=1}^{M} \frac{p_i}{\ln 2}\left[\frac{q_i}{p_i} - 1\right] = \frac{1}{\ln 2}\left[\sum_{i=1}^{M} q_i - \sum_{i=1}^{M} p_i\right]$$

$$= \frac{1}{\ln 2}\left[\sum_{i=1}^{M} q_i - 1\right] \leqq 0$$

を得る。したがって、式①が証明された。

また、等号は $q_i / p_i = 1$ ($i = 1, 2, \cdots, M$) となるとき、すなわち、$q_i = p_i$ となるとき成立する。

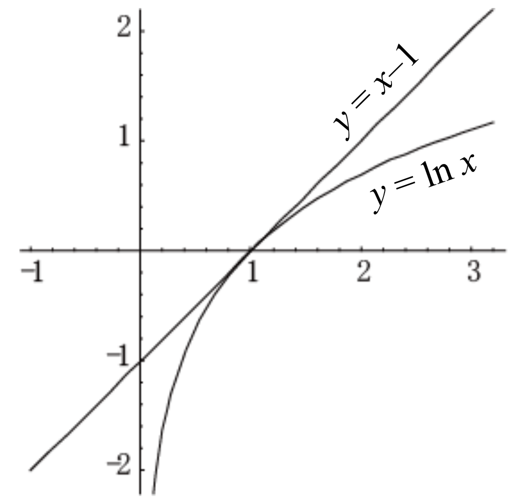


図4.3 $\ln x \leqq x-1$ の説明図

エントロピー関数

- 0と1の2つの値を確率p, 1−pの確率でとる確率変数Xのエントロピー$H(X)$は

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p)$$

これは1変数の関数

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

における$x=p$のときの値$\mathcal{H}(p)$とみることができる。

$\mathcal{H}(p)$を**エントロピー関数**とよぶ。

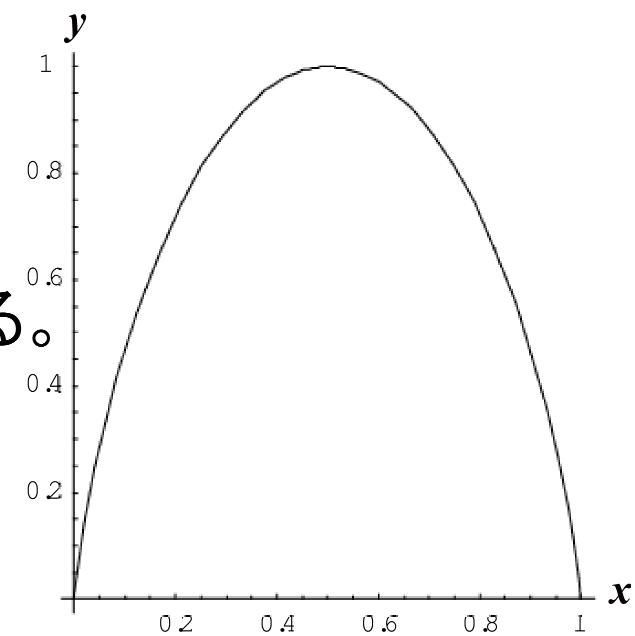
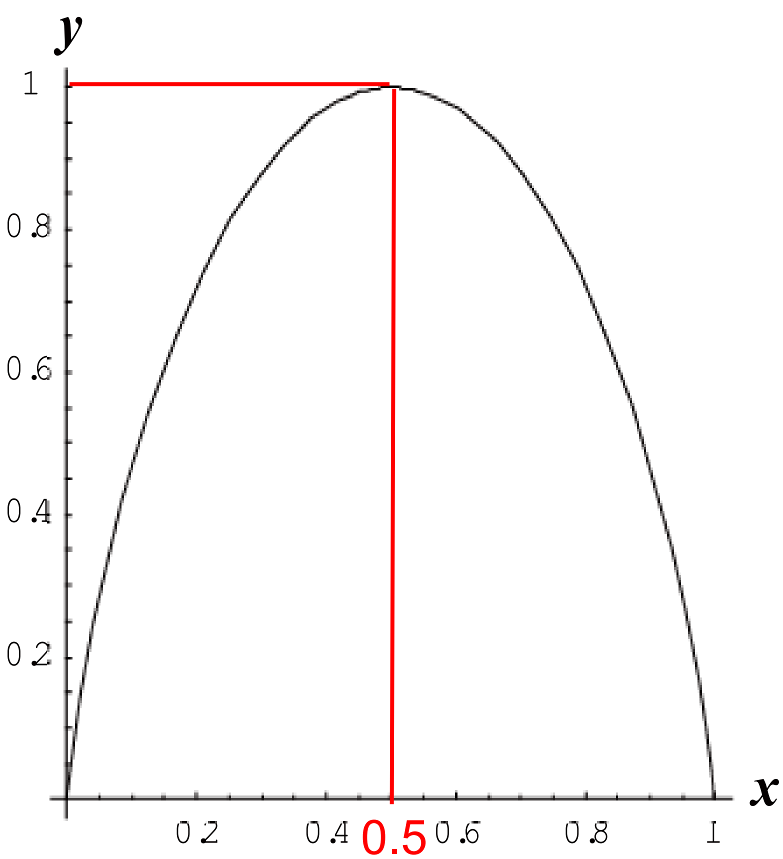


図4.8 エントロピー関数

エントロピー関数の性質



エントロピー関数

$0 \leqq \mathcal{H}(x) \leqq 1$

$x=1/2$で最大値1

$x=0, 1$で最小値0

上に凸（$d^2\mathcal{H}(x)/dx^2<0$）

直線$x=1/2$に関して対称
($\mathcal{H}(1/2+x)=\mathcal{H}(1/2-x)$)