情報システム工学科
伊達 章 (A-423)

情報理論

- 「情報とは何か」　情報を数量的にとらえることは可能なのか.

  Claude E. Shannon（クロード シャノン，天才）が，確率論を基礎に数量的に定式化した．[1]

- 例： ある病院を訪れた患者に対し，熱があったか，風邪かどうか，その割合を調べた結果

	B_1（風邪）	B_2（風邪なし）	$p(A_i)$
A_1（熱あり）	0.55	0.05	0.60
A_2（熱なし）	0.10	0.30	0.40
$p(B_j)$	0.65	0.35	

→ 熱があるかないかを知ってしまえば，風邪であるかないかはほぼ検討がつく.
→ 確率と情報量は何か関係がありそう.

- 確率分布とは

  確率と確率分布の基礎知識：

$$\begin{cases} \text{同時確率分布:} & p(A_i,B_j) & A_i \text{ と } B_j \text{ とが組になって同時に起こる確率} \\ \text{周辺確率分布:} & p(A_i) & \text{一方（}A_i\text{）だけに着目した確率分布} \\ \text{条件付確率分布:} & p(B_j|A_i) & A \text{ に関する事象が } A_i \text{ であることがわかっているときの } B_j \text{ の確率} \end{cases}$$

これらの間には次の関係が成立する（あたりまえの事を書いている．上の表で考えてみる）.

$$\sum_{i,j} p(A_i,B_j) = 1, \ \sum_i p(A_i) = \sum_j p(B_j) = 1 \tag{1}$$

$$\sum_j p(B_j|A_i) = \sum_i p(A_i|B_j) = 1 \tag{2}$$

$$p(A_i) = \sum_j p(A_i,B_j), \qquad p(B_j) = \sum_i p(A_i,B_j) \tag{3}$$

$$p(A_i,B_j) = p(A_i)p(B_j|A_i) = p(B_j)p(A_i|B_j) \tag{4}$$

$$p(B_j|A_i) = \frac{p(B_j)p(A_i|B_j)}{p(A_i)} = \frac{p(A_i,B_j)}{p(A_i)} \tag{5}$$

$$p(A_i|B_j) = \frac{p(A_i)p(B_j|A_i)}{p(B_j)} = \frac{p(A_i,B_j)}{p(B_j)} \tag{6}$$

- 計算例：

  熱がある (A_1) という条件のもとでの風邪のあるなしに関する条件付確率分布

$$p(B_1|A_1) = \frac{p(A_1,B_1)}{p(A_1)} = \frac{0.55}{0.60} = \frac{11}{12} \approx 0.92$$

$$p(B_2|A_1) = \frac{p(A_1,B_2)}{p(A_1)} = \frac{0.05}{0.60} = \frac{1}{12} \approx 0.08$$

[1] C. E. Shannon, “A mathematical theory of communication,” Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

- 情報量とエントロピー

定義：情報量

ある事象が確率 p で発生するとき，その事象が実際に起ったことを知らせる情報に含まれている情報量を $-\log_2 p$ ビットと定義する．

- 例：サイコロ ($\log_2 6 \approx 2.58$)．　コイン ($\log_2 2 = 1$)．宝くじ (-$\log_2 2^{-100} = 100$) ...
- 不確定な状況を確定したいから情報がほしい．
- どういう情報がもらえるかは事前にはわからない．
  $\rightarrow$ もらえる情報量そのものはわからない．
- $\sum_{i=1}^{n} p_i = 1$ であるので，もらえる情報量の期待値はわかる．それが次のエントロピー．

定義： エントロピー

n 個の事象が確率 $p_1, p_2, \cdots, p_n$ で発生するとき，どれが発生したかの不確定度を

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

と定義し，これをエントロピーと呼ぶ．

- この値は，不確実な状況を確定するのに要する情報量の平均．

定義： 情報量

情報を受け取る以前の，事象 A についてのエントロピーを $H(A)$，情報を受け取った後のエントロピーを $H'(A)$ とするとき，事象 A に関するこの情報の情報量 I を

$$I = H(A) - H'(A)$$

と定義する．

- 計算例：

体温を計る前の風邪についてのエントロピー：

$$H(B) = -\sum_{j=1}^{2} p(B_j) \log_2 p(B_j) = -0.65 \log 0.65 - 0.35 \log 0.35 = 0.93 \text{ ビット}$$

体温を計り，熱がある (A_1) と分かった後での風邪についてのエントロピー：

$$H(B|A_1) = -\sum_{j=1}^{2} p(B_j|A_1) \log_2 p(B_j|A_1) = -0.92 \log 0.92 - 0.08 \log 0.08 = 0.41 \text{ ビット}$$

$\Longrightarrow$ 熱があることが分かったことにより，$0.52 (= 0.93 - 0.41)$ ビットの情報量を得た！