情報理論 第 2 回
– 情報量・エントロピー (1) –
(教科書: 2 章・3 章)

野崎 隆之

今日の目的と流れ

概要

- 情報の価値や確率の不確実さを数値化したい
  $\Rightarrow$ 自己情報量・エントロピー

今日の目標

- 情報の価値を表す 自己情報量 を理解する
- エントロピーを理解し，計算できる (情報源符号化定理で利用)

今日の流れ

1. 自己情報量
2. エントロピーの定義と性質

1. 自己情報量 (1：定義と意味)

情報の持つ “価値” を数値化したい!
$\Rightarrow$ 情報 (事象) の発生確率で情報の価値を決める
(発生確率が低いほど価値が大きい)

(定義) 自己情報量

事象 a_i が発生する確率を p_i と表す.
事象 a_i が発生することで得られる自己情報量 $I(p_i)$ を次式の通りに定義する

$$I(p_i) = -\log_2 p_i \quad \text{[bit]}$$

今後，$\log_2 x$ は単に $\log x$ と書くことにする.

以降のスライドで自己情報量の持つべき 3 つの性質を列挙していき，この定義が妥当であることを示していく

1. 自己情報量 (2: 性質の導出 1)

トランプを例に説明していく

	1	2
事象	5 を引いた	ダイヤを引いた
情報量	大	小
確率	$p_1 = 4/52$	$p_2 = 13/52$

確率が小さい事象が生じたほうが情報の価値が大きい!

性質 1

$$p_1 < p_2 \Rightarrow I(p_1) > I(p_2)$$

1. 自己情報量 (3: 性質の導出 2)

[“ダイヤの 5 を引いた” という情報]
= [“ダイヤを引いた” という情報] + [“5 を引いた” という情報]

$$I[\text{“ダイヤの 5 を引いた”}] = I[\text{“ダイヤを引いた”}] + I[\text{“5 を引いた”}]$$

$$p_1 = P[\text{“ダイヤの 5 を引いた”}] = 1/52$$

$$p_2 = P[\text{“ダイヤを引いた”}] = 1/4, \qquad p_3 = P[\text{“5 を引いた”}] = 1/13$$

$$p_1 = p_2 p_3$$

性質 2

$$I(p_2 p_3) = I(p_2) + I(p_3)$$

ちょっと確率が変化したら情報量もちょっとだけ変化する

性質 3

関数 $I(p)$ は p に関して連続関数

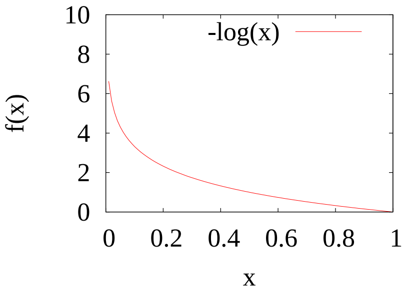
1. 自己情報量 (4: 性質のまとめ)

(まとめ) 自己情報量の持つべき性質

1. $p_1 < p_2 \Rightarrow I(p_1) > I(p_2)$
2. $I(p_2 p_3) = I(p_2) + I(p_3)$
3. $I(p)$ は p に関して連続関数

$f(x) = -K \log x$ は以上の性質を全て満たす $(K > 0)$.

1. 右図から単調減少
2. $$\begin{aligned} f(p_2 p_3) &= -K \log p_2 p_3 \\ &= -K(\log p_2 + \log p_3) \\ &= -K \log p_2 - K \log p_3 \\ &= f(p_2) + f(p_3) \end{aligned}$$
3. $\log x$ は連続関数



1. 自己情報量 (5:余談)

$f(x) = -K \log x$ は $K > 0$ に対して，情報量の性質を満たすので，
$K = 1$ の場合以外でも定義可能 (単位が変わるだけ)

- $K = 1$ の場合，
  $I(x) = -\log x$ [bit] または [シャノン (Shannon)]
- $K = 1/\log e$ の場合 (e は自然対数の底)，
  $I(x) = -\ln x$ [nat] (但し，$\ln x := \log_e x$)
- $K = 1/\log 10$ の場合，$I(x) = -\log_{10} x$ [ハートレー (Hartley)]

クロード・E・シャノン
情報理論の創始者

Wikipedia より

ラルフ・ハートレー
情報理論に貢献

Wikipedia より

1. 自己情報量 (6:余談 2)

証明はやや難しいが，実は自己情報量の性質を満たす関数は $f(x) = -K \log x$ のみである．

(定理) 自己情報量

以下の 3 つの性質を満たす関数は $f(x) = -K \log x$ のみである

1. $p_1 < p_2 \Rightarrow I(p_1) > I(p_2)$
2. $I(p_2 p_3) = I(p_2) + I(p_3)$
3. $I(p)$ は p に関して連続関数

2. エントロピー (1: 定義)

(定義) エントロピー

確率変数 X の確率分布が $p_X(x)$ で与えられたとき，エントロピー $H(X)$ は次式の通りに与えられる

$$H(X) = \sum_{x \in \mathcal{X}} p_X(x)(-\log p_X(x)) = -\sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) \quad \text{[bit]}$$

但し，$0 \log 0 = 0$ とする．

意味

自己情報量の期待値
= 一回の事象発生で得られる情報量の期待値
= 一回の観測で得られる曖昧さの減少量

2. エントロピー (2: 例)

(例) 普通のコインのエントロピー

$X(\text{表}) = 0, X(\text{裏}) = 1$

$$X : \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

$$\begin{aligned} H(X) &= -P_X(0) \log P_X(0) - P_X(1) \log P_X(1) \\ &= -1/2 \log 1/2 - 1/2 \log 1/2 = 1 \end{aligned}$$

(例) 表しかでないコインのエントロピー (予測可能な未来)

$$X : \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{aligned} H(X) &= -P_X(0) \log P_X(0) - P_X(1) \log P_X(1) \\ &= -1 \log 1 - 0 \log 0 = 0 \end{aligned}$$

エントロピーは未来の不確定さを表す量とみなせる

2. エントロピー (3:補題 1)

(補題 1) エントロピーの上限と下限

$$0 \leq H(X) \leq \log |\mathcal{X}|$$

左の不等号の等号成立条件はある $x \in \mathcal{X}$ に対して $p_X(x) = 1$.
右の不等号の等号成立条件は任意の $x \in \mathcal{X}$ に対して $p_X(x) = 1/|\mathcal{X}|$.

この補題を証明するために 2 つの補題を用意する

2. エントロピー (4:補題 2)

(補題 2)

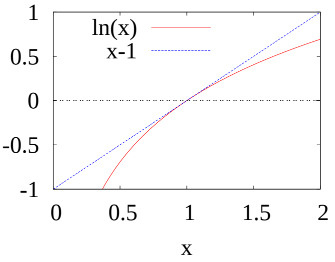
$0 < x$ について以下の不等式が成立

$$\ln x \leq x - 1$$

等号成立条件は $x = 1$ のときに限る

(証明の概略)

1. $\ln x$ が上に凸な関数であることを示す
2. 上に凸な関数では接線が関数よりも大きくなることを示す
3. $\ln x$ の $x = 1$ での接線が $x - 1$ であることを示す
4. 2,3 より，補題を得る



2. エントロピー (5:補題 3)

$|\mathcal{X}| = k$ を仮定する.

(補題 3) シャノンの補助定理

確率分布 $(p_1, p_2, \ldots, p_k)$ と $\sum_{i=1}^{k} q_i \leq 1$ を満たす $(q_1, q_2, \ldots, q_k)$ に対して，次の不等式が成り立つ

$$-\sum_{i=1}^{k} p_i \log p_i \leq -\sum_{i=1}^{k} p_i \log q_i$$

等号成立条件は全ての i に対して $p_i = q_i$.

(証明)

$$\begin{aligned}
(\text{左辺}) - (\text{右辺}) &= -\textstyle\sum_{i=1}^{k} p_i \log p_i + \sum_{i=1}^{k} p_i \log q_i \\
&= \textstyle\sum_{i=1}^{k} p_i \log \frac{q_i}{p_i} \\
&= \textstyle\frac{1}{\ln 2} \sum_{i=1}^{k} p_i \ln \frac{q_i}{p_i} \qquad (\text{底の変換公式}) \\
&\leq \textstyle\frac{1}{\ln 2} \sum_{i=1}^{k} p_i \left(\frac{q_i}{p_i} - 1\right) \qquad (\text{補題 2 より}) \\
&\leq \textstyle\frac{1}{\ln 2} \sum_{i=1}^{k} (q_i - p_i) = \frac{1}{\ln 2}\left(\sum_{i=1}^{k} q_i - \sum_{i=1}^{k} p_i\right) \leq 0
\end{aligned}$$

2. エントロピー (6: 補題 1 の証明)

$$-\sum_{i=1}^{k} p_i \log p_i = H(X) \leq \log k$$

等号成立条件は $p_1 = p_2 = \cdots = p_k = 1/k$ の時である．

(証明)：補題 3 で $q_1 = q_2 = \cdots = q_k = 1/k$ にすれば，次の不等式を得る．

$$-\sum_{i=1}^{k} p_i \log p_i \leq -\sum_{i=1}^{k} p_i \log \frac{1}{k} = \sum_{i=1}^{k} p_i \log k = \log k$$

補題 3 より，等号成立条件は全ての i に対して $p_i = q_i = 1/k$ が成立するときである．

$$0 \leq H(X) = -\sum_{i=1}^{k} p_i \log p_i$$

等号成立条件はある i に対して $p_i = 1$

(証明)：$p_i \geq 0$, $-\log p_i \geq 0$ より，$-p_i \log p_i \geq 0$ である．但し，等号成立は $p_i = 0$ または $p_i = 1$ の時に限る．
従って，$-\sum_{i=0}^{k} p_i \log p_i \geq 0$ で等号成立条件はある i に対して $p_i = 1$

今日のまとめ

自己情報量

$$I(p_i) = -\log_2 p_i \quad \text{[bit]}$$

エントロピー

$$H(X) = \sum_{x \in \mathcal{X}} p_X(x)(-\log p_X(x)) = -\sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) \quad \text{[bit]}$$