

2016年度 通信理論

第2回 情報源符号化 (情報量, エントロピー, 情報源符号化定理)

2016. 4.11

植之原 裕行 uenohara.h.aa@m.titech.ac.jp

講義スケジュール(前半)

	日付	教科書	内容
第1回	4/7	第1章・第2章	通信システム概論: 通信伝送のモデルと具体例
第2回	4/11	第3章	情報源符号化 (情報量, エントロピー, 情報源符号化定理)
第3回	4/14	第4章2節	データ圧縮 (情報源符号化, 語頭符号, ハフマン符号)
第4回	4/18	第4章3節～4節	データ圧縮(イライアス符号)
第5回	4/21	第4章5節～6節	データ圧縮(ZL符号, ランレンジス符号)
第6回	4/25	第5章	通信路容量 (通信路の確率モデル, 相互情報量, 通信路容量)
第7回	4/28	第6章1節～2節	通信路符号化 (通信路符号化, 通信路符号化定理, 繰返し符号)
第8回	5/2	第1章～ 第6章2節	通信理論に関する演習

講義スケジュール(後半)

	日付	教科書	内容
第9回	5/ 9	第6章3、4節	誤り訂正符号(線形ブロック符号)
第10回	5/12	第6章5節	誤り訂正符号(巡回符号, 畳み込み符号)
第11回	5/16	第6章6節	誤り訂正符号(LDPC符号)
第12回	5/19	第7章	連続情報と連続通信路
第13回	5/23	第8章1節～2節	暗号の基礎理論
第14回	5/26	第8章3節～7節	暗号の例(共通鍵暗号、ハッシュ関数)
第15回	5/30	第8章3節～7節	暗号の例(公開鍵暗号)
			期末試験(予定)

エントロピーと伝送容量

- エントロピーとは？

情報源の情報量の平均。

ある情報を復調可能とするための最低限のデータ量。

- 伝送容量とは？

ある通信路でエラーなしで伝送できる最大の伝送レート



本講義で定義・性質などを学ぶ

情報量

- ある離散事象 (K 個のシンボルで構成されるアルファベット) を考える。

$$\zeta = \{s_0, s_1, \dots, s_{K-1}\}$$

発生する事象が時間的に独立 → 無記憶性情報源
(Memoryless Source)

cf. マルコフ情報源: 過去の事象に
生起確率が依存

- 各シンボルの発生確率

$$P(S = s_k) = p_k \quad \text{ただし}$$

$$k = 0, 1, 2, \dots, K - 1$$

$$\sum_{k=0}^{K-1} p_k = 1$$

- 情報量 (びっくりの度合い)

$$I(s_k) = \log \left(\frac{1}{p_k} \right) \quad \rightarrow \text{底が2のとき、単位[bit]}$$

エントロピー

- すべてのシンボルに対する情報量の平均値

$$H(\zeta) = \sum_{k=0}^{K-1} p_k \log_2 \left(\frac{1}{p_k} \right) \text{ 単位: [bit]}$$

(例) $p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{8}, p_4 = \frac{1}{8}$ のシンボルのエントロピーを求める。

$$H(\zeta) = \frac{1}{2} \log_2 2 + \frac{1}{4} \log_2 4 + \frac{1}{8} \log_2 8 + \frac{1}{8} \log_2 8 = \frac{7}{4} \text{ [bit]}$$

エントロピーの性質

● 下限

$p_k = 1, p_i = 0 (i \neq k)$ のとき、 $H(\zeta) = 0$

➡ ある事象が必ず起きるならば、その情報に価値はない

なお、 $0 \leq p_k \leq 1$ なので、 $H(\zeta) \geq 0$

● 上限

$p_k = \frac{1}{K}$ のとき、 $H(\zeta) = \log_2 K$ が最大値となる。

➡ 各事象が等しい確率で発生するとき、得る情報量は最大

2元無記憶情報源のエントロピー(1)

2つのシンボル0、1がそれぞれ確率 $p_0, p_1=1-p_0$ で発生し、各シンボルの発生が独立の場合。

エントロピー $H(\zeta) = -p_0 \log_2 p_0 - (1-p_0) \log_2 (1-p_0)$

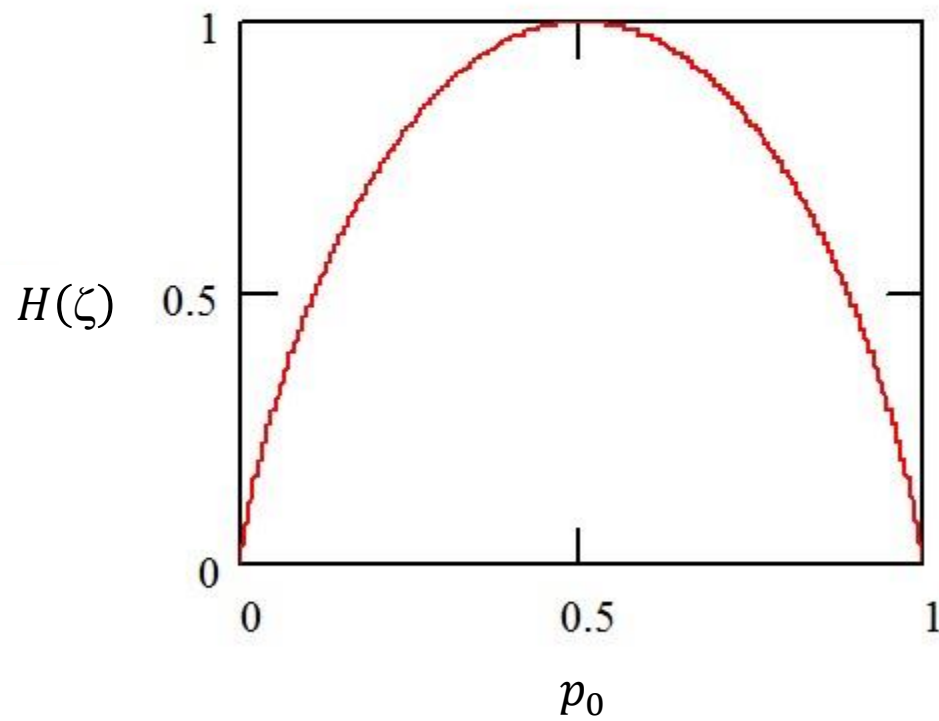
$$\begin{aligned} \frac{dH(\zeta)}{dp_0} &= -\log_2 p_0 - p_0 \frac{1}{p_0 \ln 2} + \log_2 (1-p_0) + (1-p_0) \frac{1}{(1-p_0) \ln 2} \\ &= \log_2 \frac{1-p_0}{p_0} \end{aligned}$$

$$\frac{dH(\zeta)}{dp_0} = 0 \quad \text{の条件は} \quad \frac{1-p_0}{p_0} = 1 \quad \text{なので} \quad p_0 = \frac{1}{2}$$

また $H(\zeta) = 0 \quad (p_0 = 0, 1)$

2元無記憶情報源のエントロピー(2)

計算結果



情報源の拡大とエントロピー

- 情報源の拡大 (n次拡大情報源)

ある離散事象 (アルファベット) $\zeta = \{s_0, s_1, \dots, s_{K-1}\}$ の n 個の連続する情報シンボルで構成されるアルファベットを ζ^n とする。

- 無記憶情報量

$$I(s_k, s_l, \dots, s_m) = \log \left(\frac{1}{p_k} \frac{1}{p_l} \dots \frac{1}{p_m} \right) = \log \left(\frac{1}{p_k} \right) + \log \left(\frac{1}{p_l} \right) + \dots + \log \left(\frac{1}{p_m} \right)$$

- 拡大情報源のエントロピー

$$H(\zeta^n) = nH(\zeta)$$

2次拡大情報源のエントロピー

n=2の場合: $\zeta = \{s_0, s_1, \dots, s_{K-1}\}$ に対して

$$\zeta^2 = \{s_0 s_0, s_0 s_1, \dots, s_{K-1} s_{K-1}\}$$

$$\begin{aligned} H(\zeta^2) &= \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} p_k p_l \log_2 \left(\frac{1}{p_k p_l} \right) = \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} p_k p_l \left\{ \log_2 \left(\frac{1}{p_k} \right) + \log_2 \left(\frac{1}{p_l} \right) \right\} \\ &= \sum_{l=0}^{K-1} \left(\sum_{k=0}^{K-1} p_k \right) p_l \log_2 \left(\frac{1}{p_l} \right) + \sum_{k=0}^{K-1} \left(\sum_{l=0}^{K-1} p_l \right) p_k \log_2 \left(\frac{1}{p_k} \right) \\ &= 2 \sum_{l=0}^{K-1} p_l \log_2 \left(\frac{1}{p_l} \right) \\ &= 2H(\zeta) \end{aligned}$$

情報源符号化定理

- K個のシンボルを持つアルファベットからなる情報源を考える。

k番目のシンボル s_k の発生確率: p_k ($k=0, 1, \dots, K-1$)

s_k に割り当てられた2元符号長の長さ: l_k [bits]

平均符号長 $\bar{L} = \sum_{k=0}^{K-1} p_k l_k$

- $L_{\min}$: L の最小値は、どのようにして決定されるか？



シャノンの第1定理(情報源符号化定理)

エントロピー $H(\zeta)$ の離散無記憶情報源に対して、いかなる無歪情報源の平均符号長 $\bar{L}$ は、

$$\bar{L} \geq H(\zeta)$$

まとめ

1. 情報のびっくりの度合いは、情報量で表現できる。
2. ある事象全体の情報量=エントロピー
3. 連続する n 個の情報は n 倍のエントロピーを持つ。
4. 情報源符号化定理
情報はそのエントロピーまで圧縮(符号化)可能。