2010年後期
通信伝送工学

第2回　情報源符号化

阪口 啓 sakaguchi@mobile.ee.titech.ac.jp

2010年10月12日

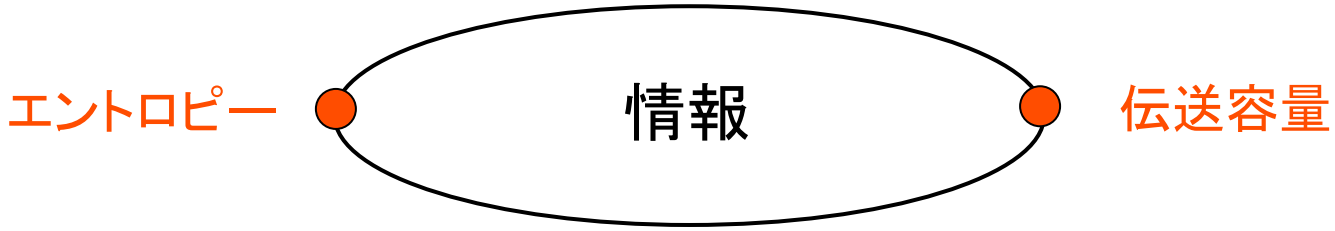
講義スケジュール前半

	日付	教科書	内容
第１回	10月5日	Preview	通信伝送のモデルと具体例
第２回	10月12日	9.1 – 9.3	情報理論の基礎１:情報源符号化
第３回	10月19日	9.4	情報理論の基礎２:データ圧縮
第４回	10月26日	9.5 – 9.7	情報理論の基礎３:通信路容量
第５回	11月2日	9.8	情報理論の基礎４:通信路符号化
第６回	11月9日	10.1 – 10.3	情報理論の基礎５:誤り訂正符号
	11月16日		中間試験

情報理論

■ エントロピー

ある情報を表現可能な最低限のデータ量



■ 伝送容量

ある通信路でエラーフリーで伝送を行える最大のデータレート

情報量

■ 離散事象（アルファベット）

$$\zeta = \{s_0 \quad s_1 \quad \cdots \quad s_{K-1}\}$$

発生する事象が時間的に独立 ⟶ 無記憶情報源

■ 発生確率

$$P(S = s_k) = p_k \qquad \text{ただし} \qquad \sum_{k=0}^{K-1} p_k = 1$$

■ 情報量（びっくりの度合い）

$$I(s_k) = \log\left(\frac{1}{p_k}\right)$$ ⟶ 低が2のとき bits

エントロピー

■ エントロピー

すべての事象に対する情報量の平均値

$$H(\zeta)=\sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

■ 例えば

$$\zeta=\{s_0 \quad s_1 \quad s_2 \quad s_3\} \qquad p_0=\frac{1}{2} \quad p_1=\frac{1}{4} \quad p_2=\frac{1}{8} \quad p_3=\frac{1}{8}$$

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

エントロピーの性質

■ エントロピーの下限

$H(\zeta)=0$　　いずれかの事象が $p_k=1$ のとき

■ エントロピーの上限

$H(\zeta)\le \log_2 K$　　事象が等確率のときにエントロピー最大

$\longleftarrow$　　基本不等式　$\displaystyle\sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right)\le 0$

$q_k = 1/K$　を代入　$\longrightarrow$　$\displaystyle\sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)\le \log_2 K$

二元無記憶情報源のエントロピー

■ 二元無記憶情報源

シンボル0の発生確率: p_0

シンボル1の発生確率: $1-p_0$

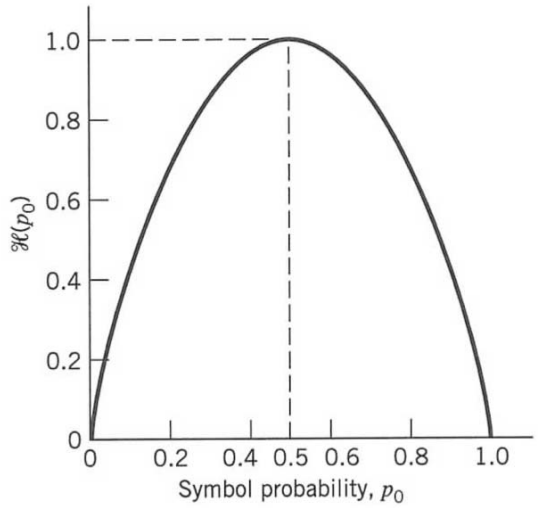


FIGURE 9.2 Entropy function $\mathcal{H}(p_0)$.

■ エントロピー（エントロピー関数）

$$H(\zeta) = -p_0 \log_2 p_0 - (1-p_0)\log_2(1-p_0)$$

$p_0 = 0$ or 1 で最小、 $p_0 = 1/2$ で最大

拡張情報源のエントロピー

■ 拡張情報源

$$\zeta^2 = \{ s_0 s_0 \quad s_0 s_1 \quad \cdots \quad s_1 s_0 \quad \cdots \quad s_{K-1} s_{K-1} \}$$

■ 拡張情報源（無記憶な場合）

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

■ 拡張エントロピー

$$H(\zeta^n) = nH(\zeta)$$ n 倍増える

情報源符号化定理

■ 平均符号長

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

l_k :事象 s_k を表現する符号の長さ

■ 情報源符号化定理

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

まで情報を圧縮することができる

■ 例えば

事象	s_0	s_1	s_2	s_3
確率	0.5	0.25	0.125	0.125
符号	0	10	110	111

⇨ エントロピー: $7/4$

平均符号長: $7/4$

まとめ

1. 情報のびっくりの度合いは情報量で表現可能
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

付録1（基本不等式）

$$\sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) - \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{q_k}\right) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right)$$

$$= \frac{1}{\log 2}\sum_{k=0}^{K-1} p_k \log\left(\frac{q_k}{p_k}\right) \le \frac{1}{\log 2}\sum_{k=0}^{K-1} p_k\left(\frac{q_k}{p_k} - 1\right)$$

$$= \frac{1}{\log 2}\sum_{k=0}^{K-1}(q_k - p_k) = 0$$

$$\longrightarrow \quad \sum_{k=0}^{K-1} p_k \log_2\left(\frac{q_k}{p_k}\right) \le 0$$

付録２（拡張情報源のエントロピー）

無記憶情報源 $P(s_k, s_l) = p_k p_l$

情報量 $$I(s_k, s_l) = \log\left(\frac{1}{p_k p_l}\right) = \log\left(\frac{1}{p_k}\right) + \log\left(\frac{1}{p_l}\right)$$

エントロピー
$$\begin{aligned} H(\zeta^2) &= \sum_{k=0}^{K-1}\sum_{l=0}^{K-1} p_k p_l I(s_k, s_l) \\ &= \sum_{k=0}^{K-1}\sum_{l=0}^{K-1} p_k p_l \log\left(\frac{1}{p_k}\right) + \sum_{k=0}^{K-1}\sum_{l=0}^{K-1} p_k p_l \log\left(\frac{1}{p_l}\right) \\ &= \sum_{k=0}^{K-1} p_k \log\left(\frac{1}{p_k}\right) + \sum_{l=0}^{K-1} p_l \log\left(\frac{1}{p_l}\right) = 2H(\zeta) \end{aligned}$$