情報理論

第7回　講義

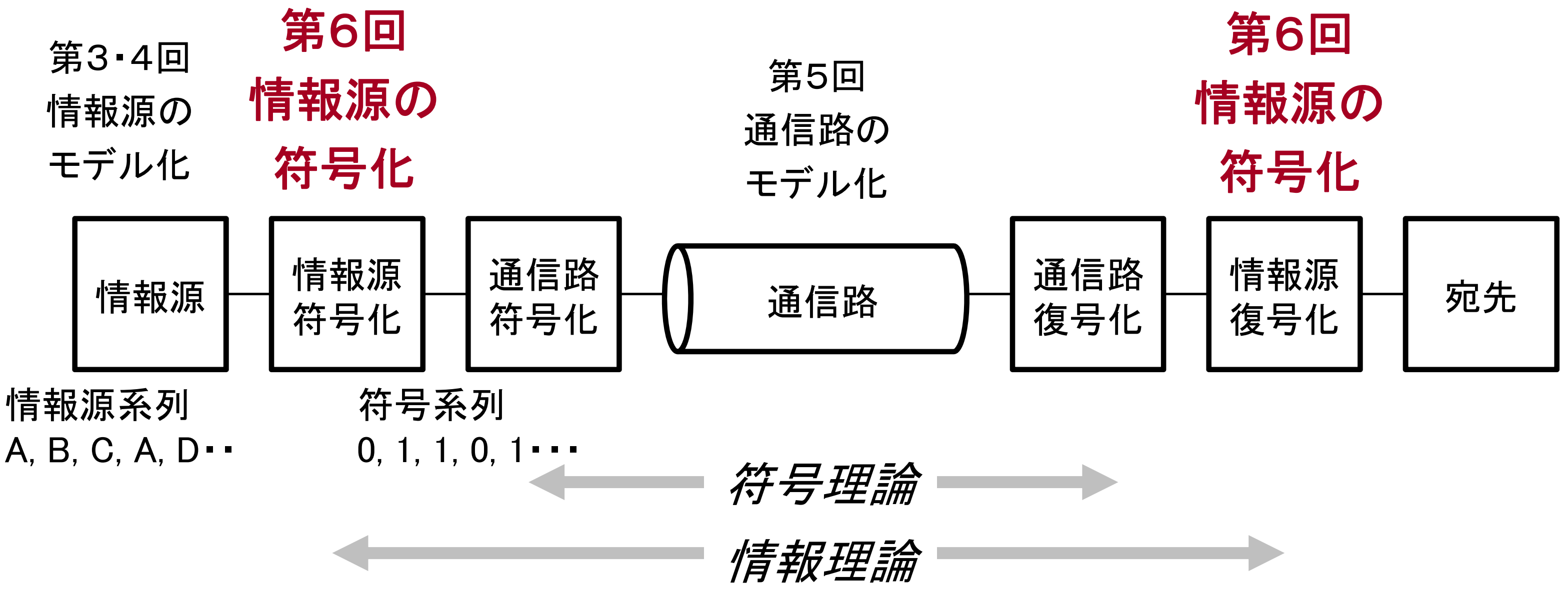
情報源の符号化(平均符号長の限界)

２０１５．６．３

植松　芳彦

本日の講義範囲と内容

1. 情報源符号化に必要な条件
2. 瞬時符号と符号の木
3. 瞬時符号と符号長の関係



クラフトの不等式

- 長さがl_1，l_2，・・・，l_MからなるM個の符号語を持つ符号が瞬時符号となる条件を考える.
- それぞれの符号語が「葉」に対応したとき，長さ l_i の符号語に行きわたる養分は 2^{-l_i}．養分の総和は1を超えない.

$$2^{-l_1} + 2^{-l_2} + \cdots + 2^{-l_M} \leq 1 \qquad \text{（式4.2）}$$

- 符号語が $\{0,1\}$ の 2 元符号でなく q 元符号の場合，個々の接点でq個に分かれるので

$$q^{-l_1} + q^{-l_2} + \cdots + q^{-l_M} \leq 1 \qquad \text{（式4.3）}$$

- 符号語の数（M ）が大きいとき，l_1，l_2，・・・，l_Mはあまり小さい値にできない.

今回やりたいこと

- クラフトの不等式が制約条件として与える場合，平均符号長の下限を導く

- 情報源 S の条件
  - 情報源アルファベット　$a_1, a_2, \cdots, a_M$
  - 各記号の発生確率　$p_1, p_2, \cdots, p_M$
- 情報源符号化 の条件
  - 一意復号可能な2元符号に符号化
  - 各符号語の長さ　$l_1, l_2, \cdots, l_M$

今回やりたいこと

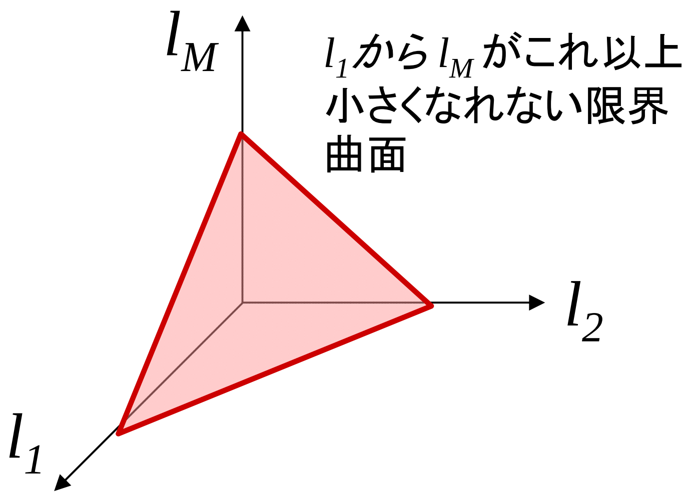
- クラフトの不等式が制約条件として与える場合，平均符号長の下限を導く

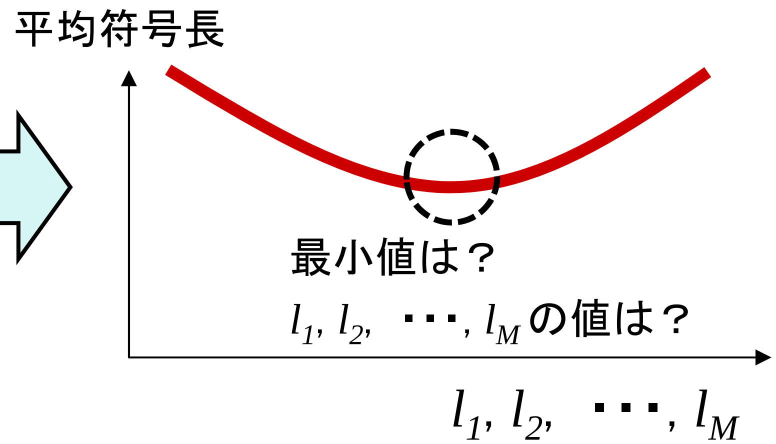
クラフトの不等式（制約条件）

$$2^{-l_1} + 2^{-l_2} + \cdots + 2^{-l_M} \quad \leq \quad 1$$

求めたいもの

$$L = l_1 \cdot p_1 + l_2 \cdot p_2 + \cdots + l_M \cdot p_M$$





やりたいことのイメージ

定理

- 平均符号長の最小値

$$H_1(S) \leq L$$ （式4.6）

- 以下の条件を満たす瞬時符号を構成可能

$$H_1(S) \leq L < H_1(S) + 1$$ （式4.7）

- 情報源Sの一次エントロピー$H_1(S)$

$$H_1(S) = -\sum_{i=1}^{M} p_i \bullet \log_2 p_i$$ （式4.8）

p は確率で1より小さいため $H_1(S)$ は必ず正
平均符号長の最小値は各符号の発生確率だけで決まる

補助定理

- p_1, ・・・, p_M (p_i は非負)に

$$p_1 + p_2 + \cdots + p_M = 1$$

- q_1, ・・・, q_M (q_i も非負)に

$$q_1 + q_2 + \cdots + q_M \leq 1 \qquad \text{（式4.9）}$$

- が成り立つとき，以下の関係が成立．

$$H_1(S) = -\sum_{i=1}^{M} p_i \bullet \log_2 p_i \leq -\sum_{i=1}^{M} p_i \bullet \log_2 q_i \qquad \text{（式4.10）}$$

等号条件　$p_i = q_i \quad (i = 1, 2, \cdots, M)$

(証明)補助定理

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

$$= -\sum_{i=1}^{M} p_i \bullet \log_2 \frac{q_i}{p_i}$$

$$= -\frac{1}{\ln 2} \sum_{i=1}^{M} p_i \bullet \ln \frac{q_i}{p_i} \quad \Longleftarrow \quad \log_b a = \frac{\log_e a}{\log_e b}$$

$$\geq \frac{1}{\ln 2} \sum_{i=1}^{M} p_i \bullet (1 - \frac{q_i}{p_i}) \quad \Longleftarrow \quad \ln x \leq x - 1, \quad -\ln x \geq 1 - x$$

（式4.11）

$$= \frac{1}{\ln 2} (\sum_{i=1}^{M} p_i - \sum_{i=1}^{M} q_i) \quad \Longleftarrow$$

単なる式展開

$$= \frac{1}{\ln 2} (1 - \sum_{i=1}^{M} q_i) \geq 0 \quad \Longleftarrow \quad \sum_{i=1}^{M} p_i = 1$$

（式4.12）

（証明）補助定理

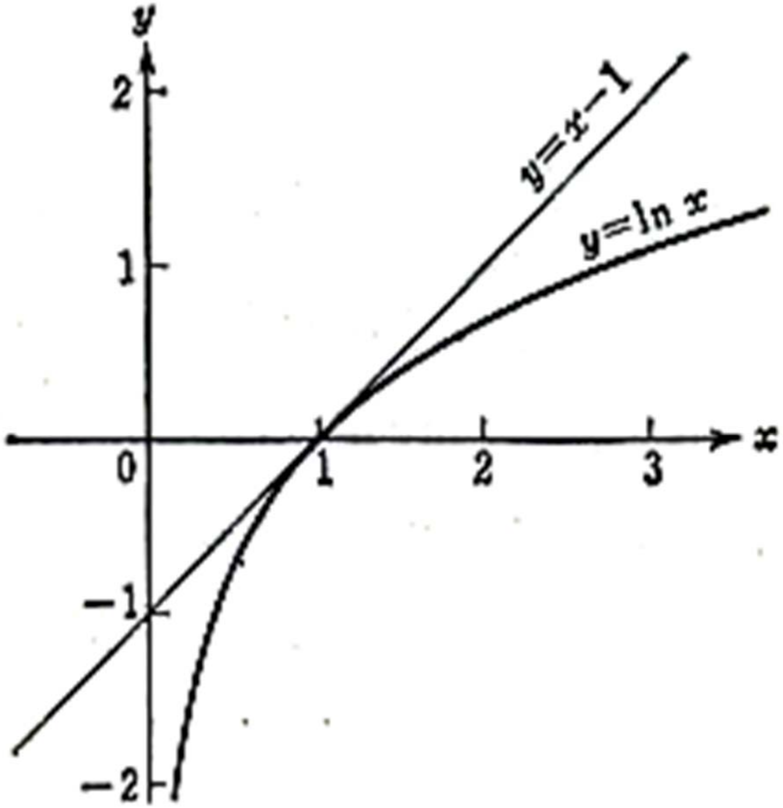


図 4.3　$\ln x \leq x-1$ の説明図

（証明）定理（1／3）

- 補助定理において， $q_i = 2^{-l_i}$ と置くとどうなるか．
- l_1, ･･･, l_M に

$$2^{-l_1} + 2^{-l2} + \cdots + 2^{-l_M} \leq 1$$

クラフトの不等式そのもの

- の関係が成り立つとき，

$$H_1(S) = -\sum_{i=1}^{M} p_i \bullet \log_2 p_i \leq -\sum_{i=1}^{M} p_i \bullet \log_2 q_i$$

$$= \sum_{i=1}^{M} p_i \bullet l_i = L$$

等号条件 $p_i = 2^{-l_i} \quad (i = 1, 2, \cdots, M)$

定理（式4.6）そのもの

（証明）定理（２／３）

- 次に実際に瞬時符号が作れることを示す，
- 以下の条件を満たす整数 $\boldsymbol{l_i}$ があることは希.

$$p_i = 2^{-l_i} \quad (i = 1, 2, \cdots, M) \qquad \text{（式4.14）}$$

- 以下の式を満たすよう整数 $\boldsymbol{l_i}$ を決めることは可能

$$-\log_2 p_i \quad \le \quad l_i \quad < \quad -\log_2 p_i + 1 \qquad \text{（式4.16）}$$

$$\frac{p_i}{2} \quad < \quad 2^{-l_i} \quad \le \quad p_i \qquad \text{（式4.17）}$$

- このように決めた $\boldsymbol{l_i}$ に対してクラフトの不等式は確かに成立する.

$$2^{-l_1} + 2^{-l2} + \cdots + 2^{-l_M} \le p_1 + p_2 + \cdots + p_M = 1 \quad \text{（式4.18）}$$

（証明）定理（３／３）

- 更に以下により定理は完全に証明される

$$-\log_2 p_i \quad \leq \quad l_i \quad < \quad -\log_2 p_i + 1$$ （式4.16）

⟸ 各辺に $\boldsymbol{p_i}$ を掛ける

$$-p_i \cdot \log_2 p_i \leq p_i \cdot l_i < -p_i \cdot \log_2 p_i + p_i$$

⟸ $\boldsymbol{i}$ =1〜M まで和をとる

$$-\sum_{i=1}^{M} p_i \cdot \log_2 p_i \leq \sum_{i=1}^{M} p_i \cdot l_i < -\sum_{i=1}^{M} p_i \cdot \log_2 p_i + 1$$

⟸ 元々の定義

$$H_1(S) \leq L < H_1(S) + 1$$

定理の実験的な証明

- エクセルシートを参照のこと.