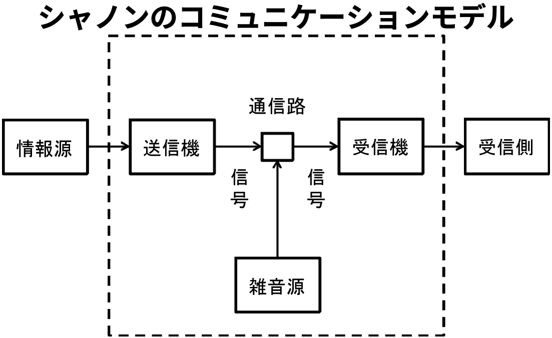
情報理論
第 6 回 情報源モデルと通信路モデル

堀田 政二
工学部 情報工学科

通信系の工学モデル (再掲)



- 情報源から発せられたメッセージは送信機で符号化され，信号として通信路に流される
- 通信路 (通信回路等) を通った信号は受信機にて復号化された後，受信側に提示される
- 情報理論で扱うのは点線部分のみ

情報伝達における情報理論の中心的課題

情報をより正確に，高速に，安く，誰にでも，いつでも，どこでも伝達するために，シャノンは情報理論の中心的課題として以下の二つを掲げている:

- 通信系において伝送すべき情報の量を明らかにすること
  - 情報源の発生情報量を明確にすること
  - 通信路の情報伝送能力を明確にすること
- 具体的な情報交換操作の考案
  - 送受信機における情報の符号化法と復号化法の考案

情報源のモデル

- デジタル情報源: 2 進数の組合せ (数字，文字，記号等)
- 情報源記号列を $S = \{s_1, s_2, ..., s_M\}$ とする．ここで s_i を情報源記号という
- s_i は数字，文字，二元情報源記号 $\{0, 1\}$ 等何でもよい

情報源のモデル化

情報源記号 s_i がどのような法則で発生するのかを，確率モデルを用いて明確にすること

記憶のない情報源と記憶のある情報源

記憶のない情報源 (memoryless source)

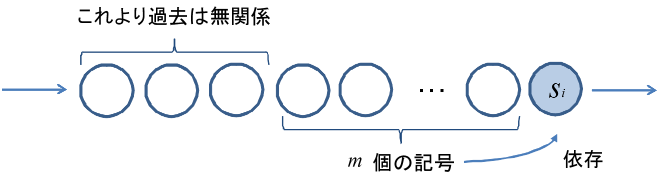
情報源記号 s_i が互いに独立に発生する情報源

- 【例】サイコロを投げる試行，コインを投げる試行

記憶のある情報源 (source with memory)

情報源記号 s_i の生起に統計的依存性がある情報源で，現在の情報源記号が過去の m 個の記号に影響を受けるもの (m 重マルコフ情報源)

- 【例】英語の文字列 (t の後は h が出やすい)



記憶のない情報源の発生情報量

情報源からどれぐらいの情報量が発生するか考えてみよう

- **記憶のない情報源のエントロピー**

M **種類の情報源記号を** s_i $(i = 1, ..., M)$**，それぞれの生起確率を** $p(s_i)$ **とする．一つの記号の持つエントロピーは**

$$H(S) = -\sum_{i=1}^{M} p(s_i) \log_2 p(s_i) \ \text{bit/記号}$$

1 **秒間に** k **個の記号が発生するとなれば，**1 **秒当たりのエントロピーは**

$$H'(S) = kH(S) \ \text{bit/sec}$$

【例】コインの表が出たら 1**，裏が出たら** 0 **を出力する情報源があるとする．一つの記号の持つエントロピーは** 1 bit**．**1 **秒間に** 10 **回コインを投げた場合，この情報源から発生する情報量は**

$$H'(S) = 10 \times 1 \ \text{bit/sec}$$

マルコフ情報源の発生情報量

- **マルコフ情報源のエントロピー**

状態 q_j **の下での** s_i **の生起確率を** $p(s_i|q_j)$ **とすると，**q_j **の下でのエントロピーは**

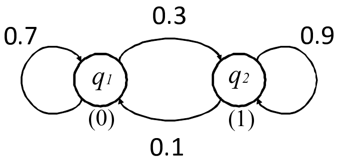
$$H(S|q_j) = -\sum_{i=1}^{M} p(s_i|q_j)\log_2 p(s_i|q_j) \ \text{bit}$$

m **重マルコフ情報源の状態の種類の数は，記号の種類が** M **で，それが** m **個ならんでいるので** $M \times M \times \cdots \times M = M^m$ **個 となる．したがって，**1 **記号あたりのエントロピーは，その平均により求められる**:

$$\begin{aligned} H(S) &= \sum_{j=1}^{M^m} p(q_j) H(S|q_j) \\ &= -\sum_{j=1}^{M^m} p(q_j) \sum_{i=1}^{M} p(s_i|q_j)\log_2 p(s_i|q_j) \\ &= -\sum_{j=1}^{M^m}\sum_{i=1}^{M} p(s_i, q_j)\log_2 p(s_i|q_j) \ \text{bit/記号} \end{aligned}$$

1重 (単純) マルコフ情報源のエントロピーの例

出力が $S = \{0, 1\}$ の単純マルコフ情報源のエントロピーを求めてみる:

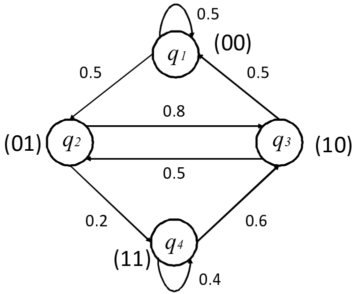


- 状態 q_1 にあるときには，0 が確率 0.7 で発生し，1 が 0.3 で発生．したがってエントロピーは
  $\mathcal{H}(0.3) = -0.7 \log_2 0.7 - 0.3 \log_2 0.3 \sim 0.881$
- 状態 q_2 にあるときには，0 を 0.1 の確率で発生し，1 を 0.9 で発生．したがってエントロピーは
  $\mathcal{H}(0.1) = -0.1 \log_2 0.1 - 0.9 \log_2 0.9 \sim 0.469$
- 一方，各状態 q_i $(i = 1, 2)$ の定常確率は，それぞれ $p(q_1) = 1/4$，$p(q_2) = 3/4$ であるので，この情報源全体のエントロピーは

$$H(S) = 1/4 \times \mathcal{H}(0.3) + 3/4 \times \mathcal{H}(0.1) \sim 0.572 \text{ bit}$$

2 重マルコフ情報源のエントロピーの例

出力が $S = \{0, 1\}$ **の 2 重マルコフ情報源のエントロピーを求めてみる**:



- **二つの記号の組合せを一つの状態と考える**
- **すると状態数は 4 つ**:

$$q_1 = (0, 0), q_2 = (0, 1), q_3 = (1, 0), q_4 = (1, 1)$$

計算結果

エルゴード的マルコフ情報源なので，遷移 (平衡) 方程式から定常確率 (分布) を求めることができる．すなわち，遷移確率行列 P，定常確率を $\boldsymbol{\alpha} = (p(q_1), p(q_2), p(q_3), p(q_4))$ とすれば

$$\boldsymbol{\alpha} = \boldsymbol{\alpha} \boldsymbol{P}$$

を解けばよく，実際に解くと

$$p(q_1) = 0.3, p(q_2) = 0.3, p(q_3) = 0.3, p(q_4) = 0.1$$

となる．したがって，この情報源全体のエントロピーは

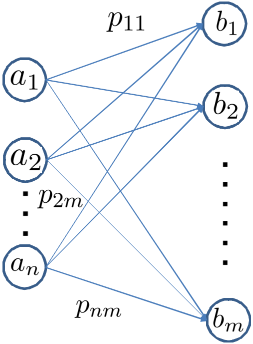
$$\begin{aligned} H(S) &= 0.3 \times \mathcal{H}(0.5) + 0.3 \times \mathcal{H}(0.2) + 0.3 \times \mathcal{H}(0.5) + 0.1 \times \mathcal{H}(0.4) \\ &\sim 0.914 \text{ bit} \end{aligned}$$

通信路のモデル

- 通信: 遠隔地どうしを結ぶ電線，電波等が必要
- 通信路: 情報理論特有の用語．通信回線のこと
- デジタル通信路: 通信路行列と通信路線図で表現できる離散通信路
- n 元通信路: 送信記号の数が n 個の $A = (a_1, a_2, ..., a_n)$ の通信路．雑音の無い理想の環境では送信記号がそのまま受信記号 $B = (b_1, b_2, ..., b_m)$ となる: $n = m$
- 実際には雑音があるので $n = m$ とは限らない (減ったり増えたりする)

通信路線図 (channel diagram)

- a_i を送信した時に b_j が受信される条件付き確率を $P(b_j|a_i) = p_{ij}$ と書き，それを辺の重みとした二部グラフを通信路線図と呼ぶ:



通信路行列 (channel matrix)

p_{ij} を並べた $n \times m$ の行列を通信路行列とよぶ:

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{pmatrix}$$

- 雑音が無いときは $n = m$ で，P は対角要素がすべて 1，すなわち，$p_{ii} = 1$ $(i = 1, ..., n)$ の $n \times n$ の単位行列となる

ベイズの定理 (再掲)

- ある事象 B と，その事象が起こる原因となる n 個の排反事象 A_1, A_2, ..., A_n を考える
- $P(A_i)$ を事前確率 (a priori probability) と呼ぶ
- $P(A_i|B)$ は B という条件下での A_i の条件付き確率で，事後確率 (a posteriori probability) と呼ぶ
- 事前確率と事後確率の関係は $P(A_i|B) = P(A_i \cap B)/P(B)$ より

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

通信路方程式

結合確率 $P(a_i \cap b_j)$ は

$$P(a_i \cap b_j) = P(a_i)P(b_j|a_i) = P(b_j)P(a_i|b_j)$$

より

$$P(a_i|b_j) = P(a_i \cap b_j)/P(b_j)$$

となる．これは，b_j を受信したという条件のもとでの送信記号 a_i の事後確率 (発生確率)．ここで

$$P(b_j) = \sum_{i=1}^{n} P(a_i)P(b_j|a_i)$$

であり，これを通信路方程式と呼ぶ．この通信路方程式を用いると，事後確率 $P(a_i|b_j)$ は

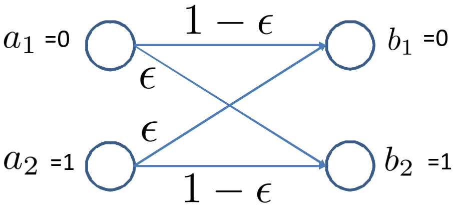
$$P(a_i|b_j) = P(a_i \cap b_j)/\sum_{i=1}^{n} P(a_i)P(b_j|a_i)$$

となる．これにより b_j を受信したとき，元の送信記号が a_i である条件付き確率は，送信記号の発生確率 $P(a_i)$ と通信路行列 p_{ij} から求められることが分かる

二元対称通信路 (BSC, Binary Symmetric Channel)

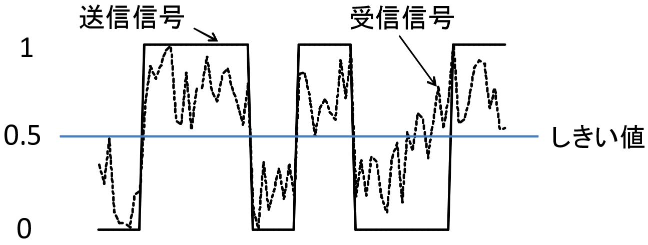
次のような通信路行列 P で定義される二元通信路のこと

$$\boldsymbol{P} = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$



- $0 \leq \epsilon < 1/2$ **は雑音による誤り発生率で，誤り率と呼ばれる**
- ϵ **の具体的な値は通信路によって異なるが，おおよそ** $10^{-1} \sim 10^{-10}$ **程度の値**
- 2 **元を** n **元に一般化したものを** n **元対称通信路と呼ぶ** (nSC)

BSC における受信記号の判定例



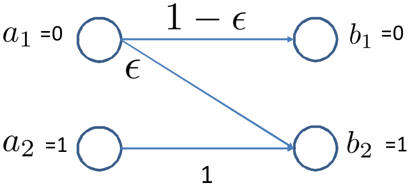
送信記号	0	1	1	0	1	0	0	1
受信記号	0	1	1	0	1	0	1	1

- **送信信号は二値パルス波．受信信号には雑音が加わる**
- **しきい値以上の受信信号を 1，それ以外を 0**

二元非対称通信路 (BAC, Binary Asymmetric Channel)

次のような通信路行列 P で定義される二元通信路のこと

$$\boldsymbol{P} = \begin{pmatrix} 1-\epsilon & \epsilon \\ 0 & 1 \end{pmatrix}$$



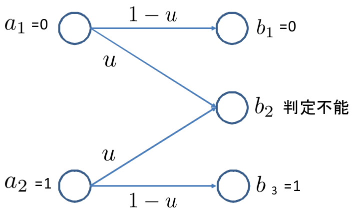
- 1 は常に正しく 1 と受信できるが，0 は確率 ϵ で 1 に誤る可能性のある通信路
- 半導体メモリ: 荷電が放電して $1 \to 0$ になる
- 光通信: 発光 $\to 1$， 消滅 $\to 0$，

二元消失通信路 (BEC, Binary Erasure Channel)

- **消失通信路: 判定不能を設定した通信路**
- **受信記号がどれにあたるのか，わからない状態**
- **信号が弱かったり，雑音がひどかったりするような場合には有効**

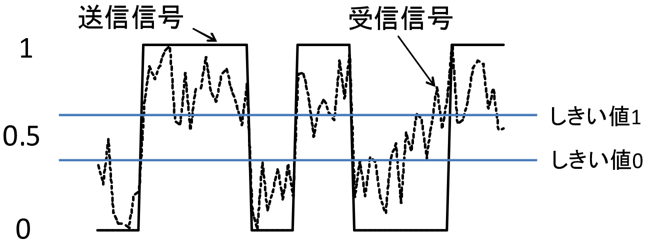
通信路行列 P **は，判定不能率** $0 \le u \le 1$ **を使って次のように定義できる:**

$$
\boldsymbol{P} = \begin{pmatrix} 1-u & u & 0 \\ 0 & u & 1-u \end{pmatrix}
$$



ただし $0 \to 1$，$1 \to 0$ **という誤りはないものと仮定**

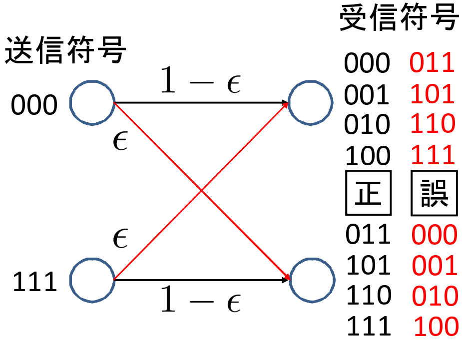
BEC における受信記号の判定例



送信記号	0	1	1	0	1	0	0	1
受信記号	0	1	1	0	1	0	?	1

- **二つのしきい値の間にある判定しにくい範囲を消失と判定**

記号 0 **を** 000**，記号** 1 **を** 111 **と符号化し，復号化は多数決により行うとする**:



一つの符号が正しく伝送される確率は

$$(3\text{ ビットとも正しい確率}) + (2\text{ ビット正しい確率}) \times 3$$

$$= (1-\epsilon)^3 + 3(1-\epsilon)^2\epsilon$$

となる．平均誤り率 P_E **は** 0 **の生起確率を** p**，**1 **の生起確率を** $1-p$ **とすれば以下のように計算できる**:

$$P_E = 1 - (0\text{ の生起確率}) \times (\text{符号が正しく伝送される確率}) - (1\text{ の生起確率}) \times (\text{符号が正しく伝送される確率})$$

$$= 1 - (1-\epsilon)^3 - 3(1-\epsilon)^2\epsilon = 3\epsilon^2 - 2\epsilon^3$$

問題

- 【6.1】次の遷移確率行列で表される単純マルコフ情報源のエントロピーを求めよ:

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 0.6 & 0.4 \end{pmatrix}$$

  ただし $\mathcal{H}(0.6) = \mathcal{H}(0.4) \sim 0.971$ としてよい．
- 【6.2】二元消失通信路で $0 \to 1$，$1 \to 0$ の誤りのある場合の通信路行列 P と通信路線図を描け．ただし，消失確率を u，誤り率を ϵ とする．