情報理論　#1

第1回講義　4月18日

情報科学研究科　　井上純一

http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/

情報量とは何か？

情報量：　$-\log p$　　$p(E)=p$　はある事象Eが起こる確率

抽象的概念である「情報」の量の数学的な定義

例1)　明日の天気に関し、p(晴れ)=p(雨)=1/2の場合、「明日は晴れ」という通報を受けた場合に得られる情報量

$$-\log p(\text{晴れ}) = \log 2 = 1(\text{ビット})$$

例2)　「犬が人間に噛みつく確率」p(犬→人間)=2^{-3}、「人間が犬に噛みつく確率」p(人間→犬)=2^{-1000}

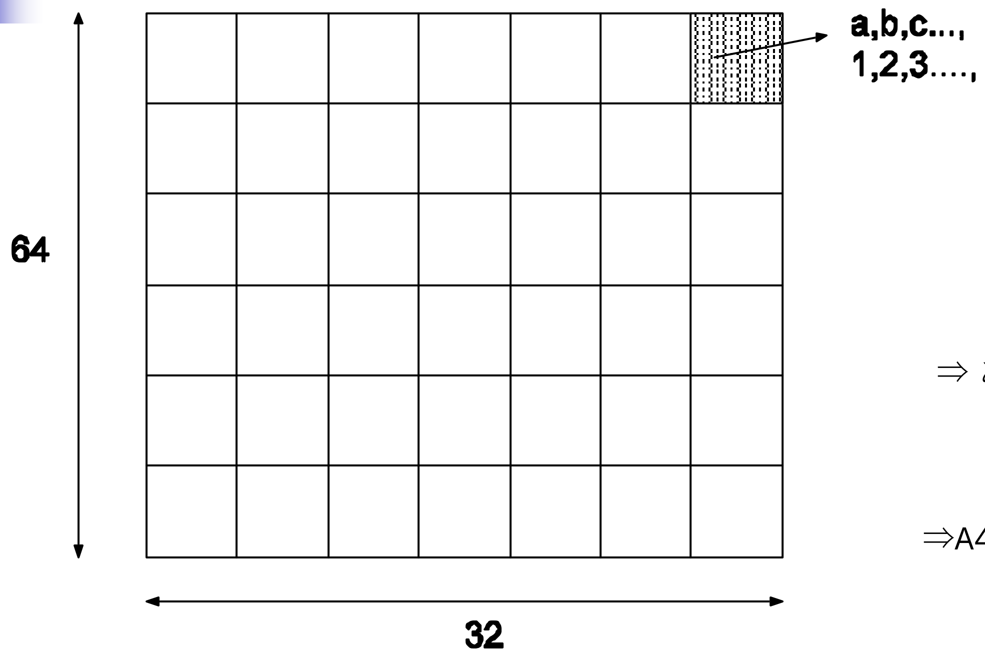
「犬が人間に噛みついた」という通報を受けた際に得られる情報量　$-\log p(\text{犬} \to \text{人間}) = 3(\text{ビット})$

「人間が犬に噛みついた」という通報を受けた際に得られる情報量　$-\log(\text{人間} \to \text{犬}) = 1000(\text{ビット})$

めったに起こらない事象に対する情報量は大きい

⇒　情報量とはある通報で我々が驚く度合いを表している

A4用紙1ページ分の情報量を見積もる



一マスには

$$2^{12} = 4096$$

種類の記号が書き込める

⇒ある1マスにある記号が書き込まれる確率

$$p = 2^{-12}$$

⇒ ある一マスにある文字を見て我々が得る情報量

$$-\log p = 12(\text{ビット})$$

⇒A4用紙1ページでは

$$12 \times 64 \times 32 = 12 \times 2^{11} (\text{ビット})$$

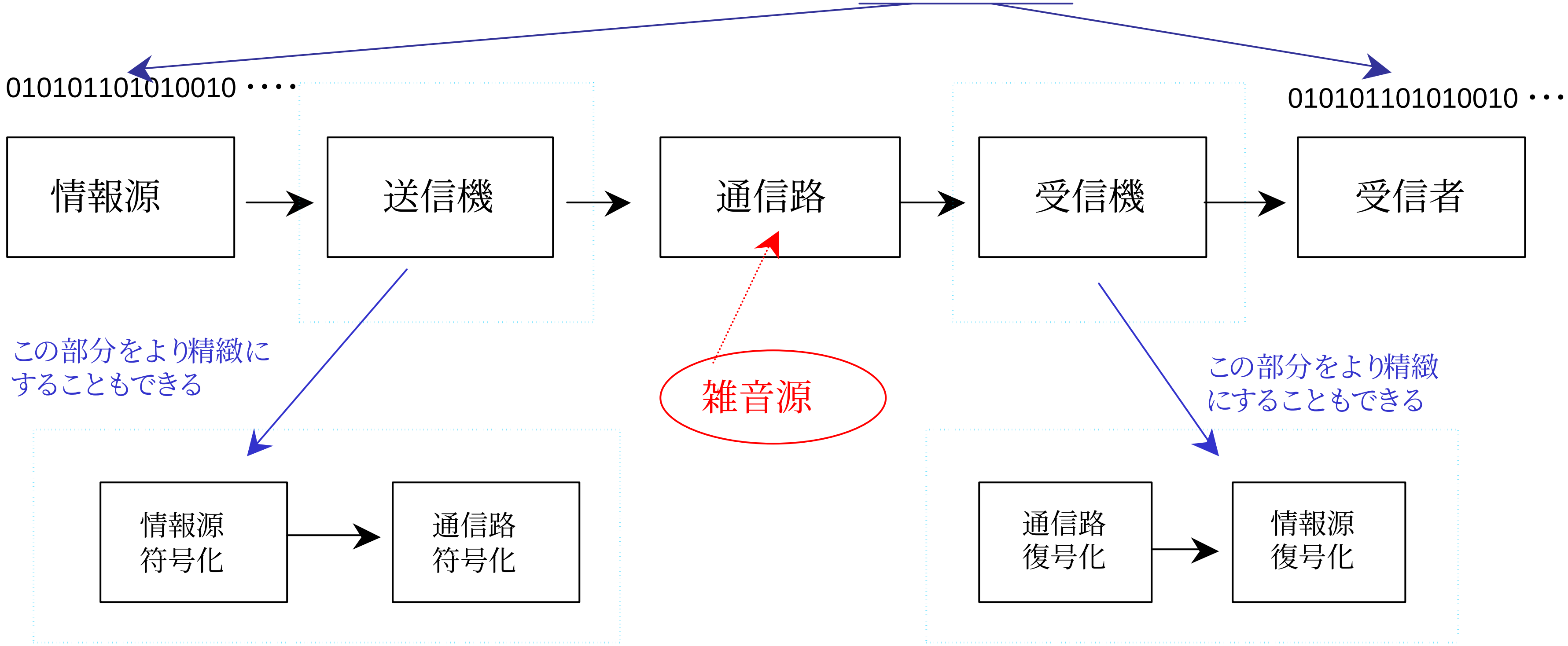
8 (ビット) ⇒ 1(バイト)、8 x 2^{10} (ビット) ⇒ 1(キロバイト, KB)と単位を約束すると

A4用紙1ページ分の情報量は

$$\boxed{\frac{12 \times 2^{11}}{8 \times 2^{10}} = 15(\text{KB})}$$

情報通信路のモデル

アルファベット、記号、数字、漢字 ⇒ 0,1からなる**デジタル情報**で表現される



符号化（情報の圧縮・冗長性の付加）と**復号化（もとの情報を復元すること）**は効率がよく、安全な情報通信にとって不可欠である。

情報源の確率モデル

離散的M 元情報源　:決められた時間ごとにM種類の記号の中から逐次生成させる情報源

記号列　$s_0, s_1, \ldots, s_n$　が発生したとすると、その**生成確率**は次のような**確率**で表現できる

$$\boxed{P(X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) = P(s_0, s_1, \ldots, s_n)}$$

結合分布

時刻　n　に情報源が発生した記号を表す確率変数

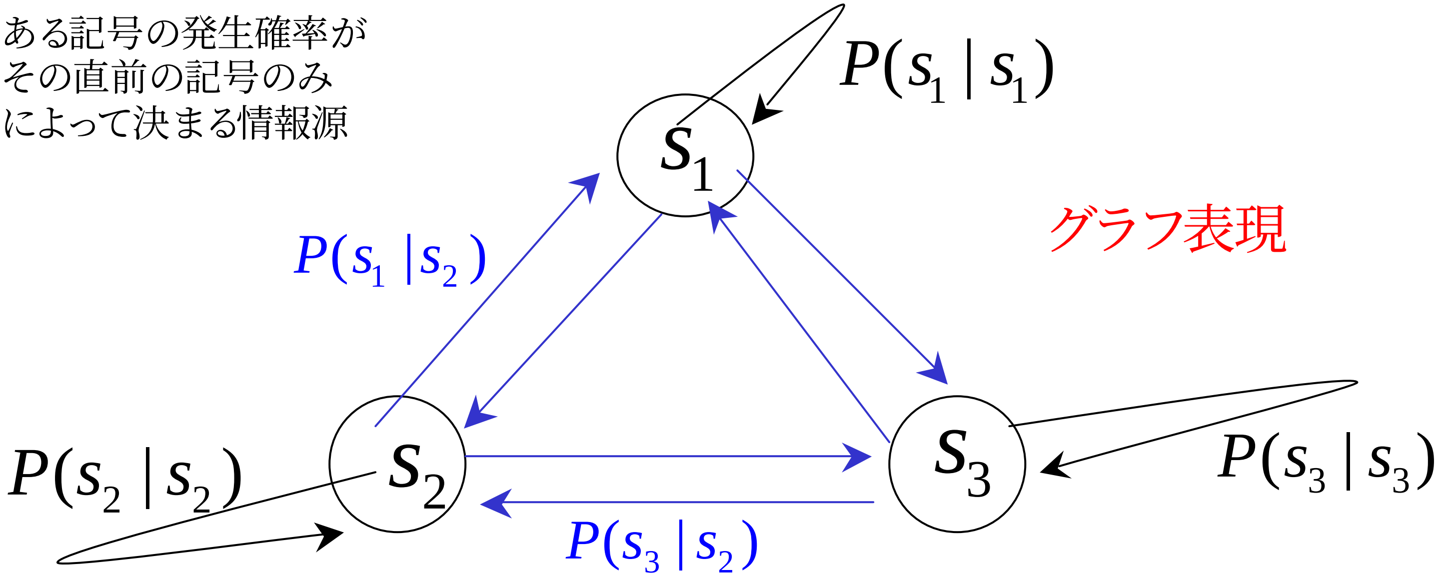
具体的に理論を構築する際には次のような単純化をおこなう

記憶の無い情報源　: $P(X_i, X_j) = P(X_i)P(X_j)$　全ての確率変数が独立である

定常情報源　:　$P(X_i = s) = P(X_j = s) = P(s)$　各時刻において、記号の発生確率が等しい

定常的単純マルコフ情報源

ある記号の発生確率が
その直前の記号のみ
によって決まる情報源



グラフ表現

行列表現

$$\mathbf{P} = \begin{pmatrix} P(s_1 \mid s_1) & P(s_1 \mid s_2) & P(s_1 \mid s_3) \\ P(s_2 \mid s_1) & P(s_2 \mid s_2) & P(s_2 \mid s_3) \\ P(s_3 \mid s_1) & P(s_3 \mid s_2) & P(s_3 \mid s_3) \end{pmatrix}$$

遷移行列

通信路の確率モデル

系列　：$s_0, s_1, \ldots, s_n$　を送信し、系列　：$t_0, t_1, \ldots, t_n$　を受信したとすると

$$\boxed{P(Y_0 = t_0, Y_1 = t_1, \ldots, Y_n = t_n \mid X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n)}$$

が通信路の性質を決める

具体的に理論を構築する際には次のような簡略化を行う

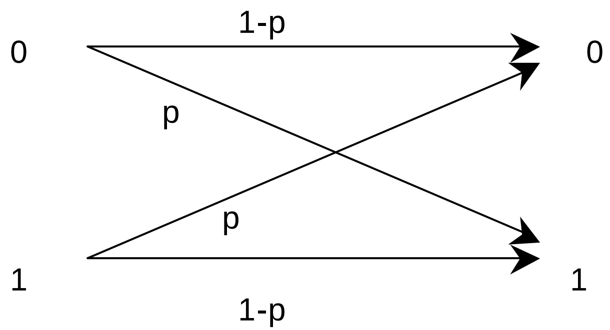
定常無記憶通信路　：

$$P(Y_0 = t_0, Y_1 = t_1, \ldots, Y_n = t_n \mid X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n)$$

$$= \prod_{i=1}^{n} P(Y_i = t_i \mid X_i = s_i)$$

各時刻で入出力関係が独立である

2元対称通信路



$$\mathbf{P} = \begin{pmatrix} P(0|0) & P(0|1) \\ P(1|0) & P(1|1) \end{pmatrix} = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

誤り確率

$P(0) = P(1) = 1/2$　のように情報源の確率モデルを選ぶと

$$P_0 = \sum_{i=0,1} P(0|i)P(i) = P(0|0)P(0) + P(0|1)P(1) = 1/2$$

$$P_1 = \sum_{i=0,1} P(1|i)P(i) = P(1|0)P(0) + P(1|1)P(1) = 1/2$$

例題 1、演習問題 1を参照

の確率で0，1が出力される