情報理論　#7

第7回講義　6月6日

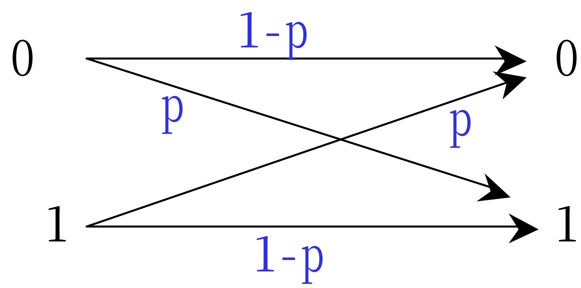
情報科学研究科　　井上純一

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

雑音のある状況でのデータ送信

誤り率 p の2元対称通信路

ビット反転率



ただ1度だけでなく、何回か同じ記号を繰り返し送信する

受信者側は多数決で送信された記号を確定する

5回送信した場合：00101 ⇒ 0、 01111 ⇒ 1 等

送信回数 n = 5 のときの誤り確率は

$$f_e^{(5)}(p) = {}_5C_3\, p^3(1-p)^2 + {}_5C_4\, p^4(1-p) + p^5$$
$$= 10p^3(1-p)^2 + 5p^4(1-p) + p^5$$

3ビットの誤りが5ビットの何番目にくるかという場合の数

誤り確率の繰り返し回数依存性

$$f_e^{(3)}(p) = 3p^2(1-p) + p^3$$

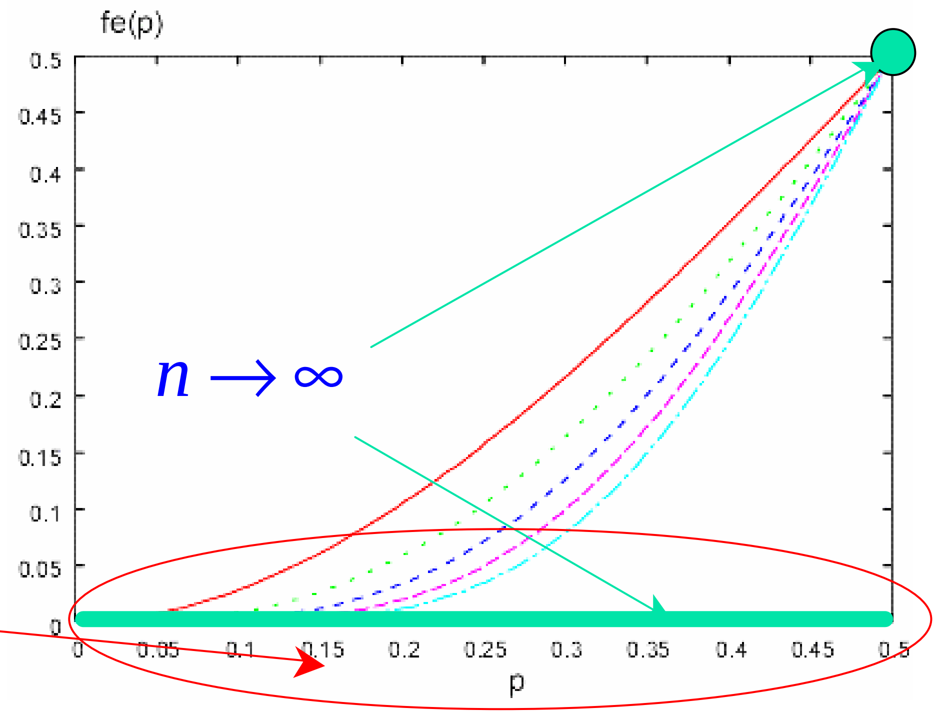
$$f_e^{(5)}(p) = 10p^3(1-p)^2 + 5p^4(1-p) + p^5$$

$$f_e^{(7)}(p) = 35p^4(1-p)^3 + 21p^5(1-p)^2 + 7p^6(1-p) + p^7$$ のように振舞う

$n \to \infty$ではどうなるか？

$$f_e^{(\infty)} = \begin{cases} 0 & (0 \le p < 1/2) \\ \dfrac{1}{2} & (p = 1/2) \end{cases}$$

導出の詳細は講義ノート参照



送信回数を増やせば、
誤り確率はいくらでもゼロに近づく

通信路容量

$$C = \max_{P_X} I(X;Y)$$

$$I(X;Y) = H(Y) - H(Y \mid X)$$

通信路容量の定義

X : 入力
Y : 出力

相互情報量 : 入力Xを知ったときに、出力Yに対して得られる知識の増加分

入力の確率分布に関して、相互情報量を最大化したものが通信路容量

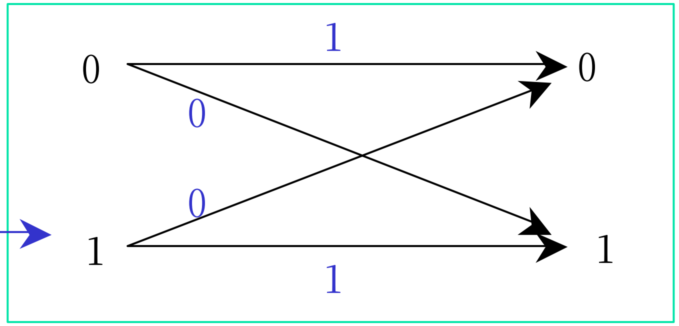
通信路容量は、通信路が伝送できる最大の情報量という意味を持つ

通信路容量の計算例

誤りの無い2元対称通信路

$$P_{Y|X}(0|0) = P_{Y|X}(1|1) = 1$$
$$P_{Y|X}(0|1) = P_{Y|X}(1|0) = 0$$

条件付き確率での表現



グラフ表現

$P_X(0) = p,\ P_X(1) = 1 - p$ のように入力分布をおくと、出力分布は

$$P_Y(0) = \sum_{x=0,1} P_{Y|X}(0|x) P_X(x) = p,\quad P_Y(1) = \sum_{x=0,1} P_{Y|X}(1|x) P_X(x) = 1 - p$$

$$I(X;Y) = H(Y) - H(Y|X) = h(p)$$
$$= \boxed{-p\log p - (1-p)\log(1-p)}$$

2値エントロピー関数
p= 1/2で最大値 1をとる

通信路容量は $C = \max_p I(X;Y) = h(1/2) = 1$

多次元入出力-定常無記憶通信路の相互情報量の上限 と通信路容量の関係#1

$$\mathbf{X} = X_1, X_2, \ldots, X_n \in A^n,\ \mathbf{Y} = Y_1, Y_2, \ldots, Y_n$$

入力ベクトル　出力ベクトル

出力の同時分布に関し

$$P_Y(Y_1, Y_2, \ldots, Y_n) = P(Y_n \mid Y_{n-1}, \ldots, Y_1) P(Y_{n-1}, \ldots, Y_1)$$

$$P(Y_{n-1}, \ldots, Y_1) = P(Y_{n-1} \mid Y_{n-2}, \ldots, Y_1) P(Y_{n-2}, \ldots, Y_1)$$

確率の積の公式

....

$$P(Y_3, Y_2, Y_1) = P(Y_3 \mid Y_2, Y_1) P(Y_2, Y_1)$$

$$P(Y_2, Y_1) = P(Y_2 \mid Y_1) P(Y_1)$$

従って

$$\boxed{P_Y(Y_1, Y_2, \ldots, Y_n) = P(Y_n \mid Y_{n-1}, \ldots, Y_1) P(Y_{n-1} \mid Y_{n-2}, \ldots, Y_1) \cdots P(Y_1)}$$

が成り立つ

多次元入出力-定常無記憶通信路の相互情報量の上限 と通信路容量の関係#2

$$P_Y(Y_1, Y_2, \ldots, Y_n) = P(Y_n \mid Y_{n-1}, \ldots, Y_1) P(Y_{n-1} \mid Y_{n-2}, \ldots, Y_1) \cdots P(Y_1)$$ より

$$H(\mathbf{Y}) = -\sum_{Y_1} \cdots \sum_{Y_n} P_Y(Y_1, \ldots, Y_n) \log P_Y(Y_1, \ldots, Y_n) \boxed{= \sum_{i=1}^{n} H(Y_i \mid Y_1, \ldots, Y_i)}$$

これを示せというのが、今週の演習問題の一つ

無記憶な通信路を考えているので

$$P_{Y|X}(\mathbf{Y} \mid \mathbf{X}) = \prod_{i=1}^{n} \boxed{P_{Y|X}(Y_i \mid X_i)}$$

各時刻で「誤り方」が独立

条件付きエントロピーは

$$H(\mathbf{Y} \mid \mathbf{X}) = -\sum_{\mathbf{X}} \sum_{\mathbf{Y}} P_{XY}(\mathbf{X}, \mathbf{Y}) \log \prod_{i=1}^{n} P_{Y|X}(Y_i \mid X_i)$$

$$= \sum_{i=1}^{n} \left\{ -\sum_{\mathbf{X}} \sum_{\mathbf{Y}} P_{XY}(\mathbf{X}, \mathbf{Y}) \log P_{Y|X}(Y_i \mid X_i) \right\} = \sum_{i=1}^{n} H(Y_i \mid X_i)$$

多次元入出力-定常無記憶通信路の相互情報量の上限 と通信路容量の関係#3

相互情報量は

$$I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n} H(Y_i \mid Y_1,\ldots,Y_n) - \sum_{i=1}^{n} H(Y_i \mid X_i)$$

$H(Y_i \mid Y_1,\ldots,Y_{i-1}) \le H(Y_i)$

条件を増やせば
あいまいさは減少する

$$\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) = \sum_{i=1}^{n} I(X_i;Y_i) \le nC$$

通信路容量の定義

$$\boxed{I(\mathbf{X};\mathbf{Y}) \le nC}$$

伝送速度と誤り確率

$$R = \frac{\log M}{n} = \frac{1}{n}$$

送信記号の種類の数
1, 0ならば 2

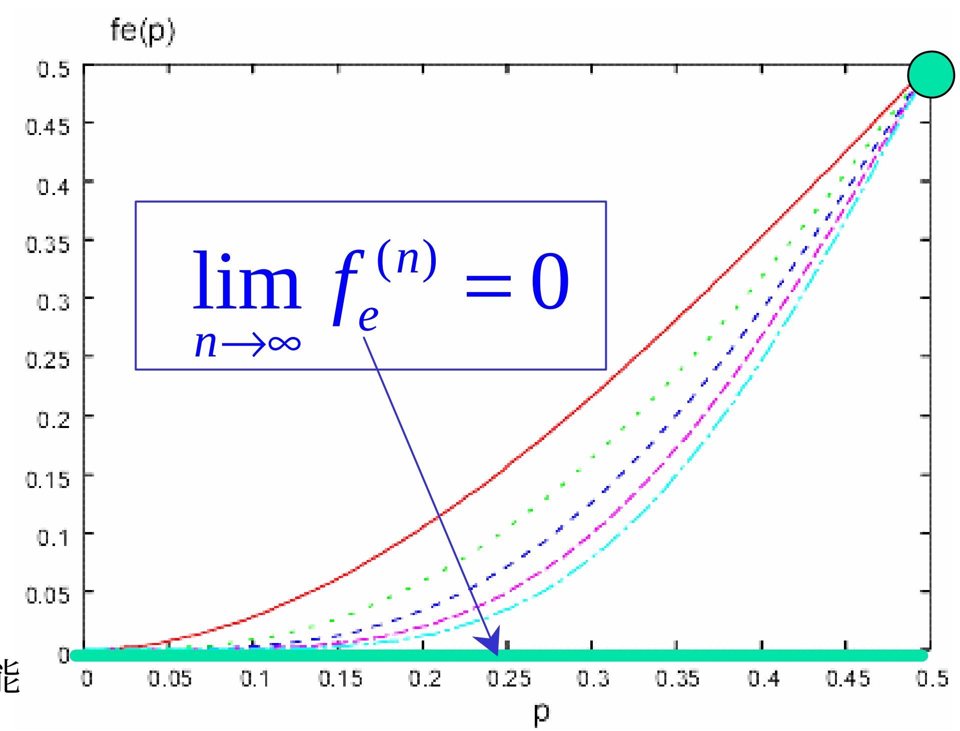
伝送速度（レート）

繰り返し回数を限りなく多くとると、
誤り確率は限りなくゼロに近づくが、
同時に伝送速度もゼロになってしまう

⇒　実用的ではない。しかし

$$R < C$$

ならば、限りなく小さな誤り確率での情報伝送が可能



$$\lim_{n\to\infty} f_e^{(n)} = 0$$

詳細は次回