情報理論　#8

第8回講義　6月13日

情報科学研究科　　井上純一

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

伝送速度と誤り確率

前回の復習）

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

送信記号の種類の数
1，0ならば2

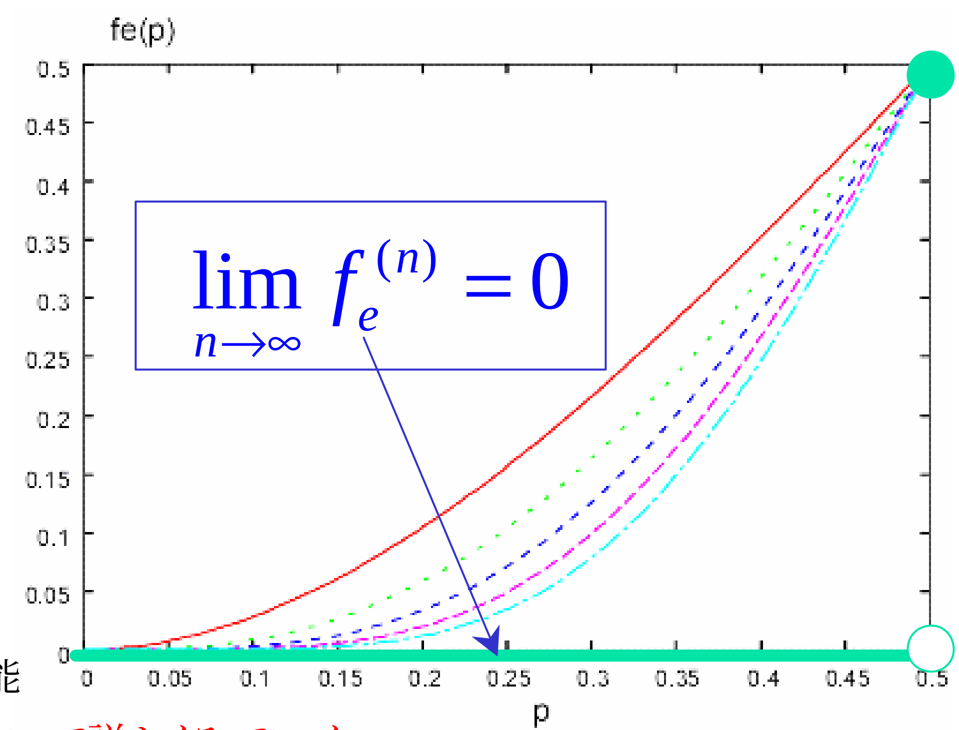
伝送速度（レート）

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



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

今回はこの事実：通信路符号化定理について詳しくみていく

通信路符号化定理

通信路符号化定理

i) $R < C$なる任意の速度$R = {}^{\log M}\!/_{n}$に対し

任意に小さな復号誤り率p_Eの符号が存在する。

ii) $R > C$なるRに対し、任意に小さなp_Eの符号は存在しない。

ゼロでなくてもよい。しかし、符号長 n を無限大ととることは必要

情報源の記号数 $M = 2^{nR}$

この定理を「ランダム符号」と呼ばれる符号に対して証明していく

ランダム符号

情報源の記号: $S_1, S_2, \ldots, S_M$　**（等確率で生成される）**

符号化　：

$S_1, S_2, \ldots, S_M$ **の各々に****ランダムに0,1をn個並べた系列****を割りあてる**

$$S_1 \leftarrow 0010 \cdots 000$$

$$S_2 \leftarrow 0100 \cdots 001$$

$$S_3 \leftarrow 0010 \cdots 100$$

$$\cdots\cdots\cdots$$

$$S_M \leftarrow \boxed{1000 \cdots 010}$$

符号長 n

各ビットに 1/2 の確率で 0,1 を割り振る

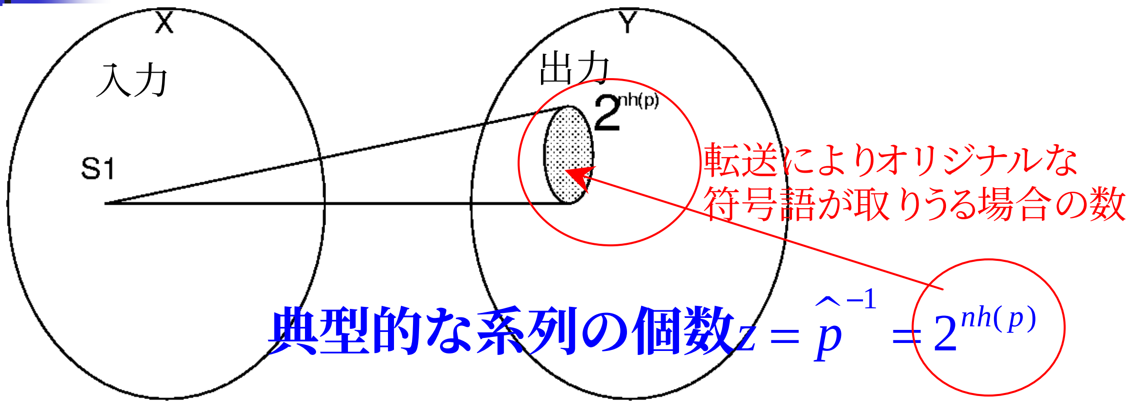
$M = 2^{nR} < 2^{n}$

$R \leq 1$　**として議論を進める**

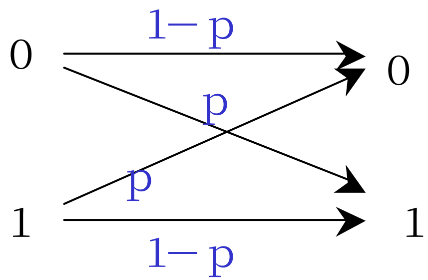
これらを2元対称通信路を介して転送する状況下で定理を証明していく

通信路符号化定理の証明#1

一つの符号語が転送により拡大する大きさの評価



典型的な系列の個数 $z = \hat{p}^{-1} = 2^{nh(p)}$



符号語は2元対称通信路
を介して転送される

ある符号語$S_1 = 000\cdots000$**は通信路の雑音により**$010\cdots101$**として受信される**

全n**ビットのなかで誤りの生じるビット数はおおよそ**np**と見積もられる**

$$\underbrace{(0100\cdots011)}_{np\ \text{ビットの誤り}},\ \underbrace{(0111\cdots100)}_{np\ \text{ビットの誤り}},\ \cdots,\ \underbrace{(0101\cdots111)}_{np\ \text{ビットの誤り}}$$

通信路出力の
典型的な系列

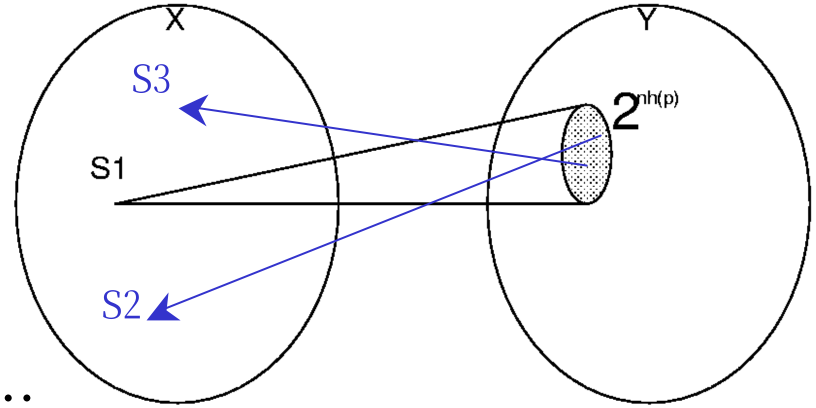
典型的な系列のなかの一つが現れる確率

$$\hat{p} = p^{np} \cdot (1-p)^{n(1-p)} = 2^{np\log n + n(1-p)\log(1-p)} = 2^{-nh(p)}$$

2値エントロピー関数

通信路符号化定理の証明#2

復号誤り率の評価



復号誤りが生じるのは・・・

受信される可能性のある$2^{nh(p)}$個の各々が、実際に送信された符号語S_1以外の$S_2, S_3, \ldots, S_M$の$(M-1)$個のいずれかに間違って復号されるとき

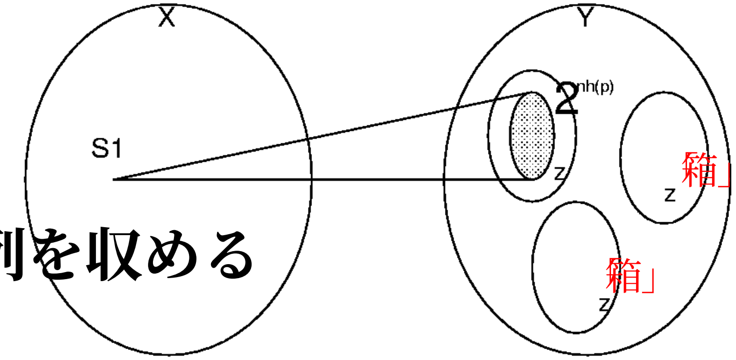
$$p_E = \underbrace{\frac{(M-1)}{2^n}}_{\text{(M-1)個のいずれかが選ばれる確率}} \times \underbrace{2^{nh(p)}}_{\text{受信される可能性のある系列の個数}} \simeq \frac{M}{2^n} \cdot 2^{nh(p)} = 2^{n(R-1+h(p))} = 2^{n(R-C)}$$

$\rightarrow 0 \quad (C = 1 - h(p) > R)$

i) の証明終わり

通信路符号化定理の証明#3

ii)の証明



受信系列の全ての可能性2^n

1つの符号語に対する通信路出力の典型列を収める

「箱」のサイズ$= z = 2^n/M$

2元対称通信路により実際に得られる典型列の数$2^{nh(p)}$

は全てこの「箱」に収まらなければならない

$$\frac{2^n}{M} > 2^{nh(p)} \Rightarrow 2^{n(C-R)} > 1$$

これは$R > C$では満たされない

ii)の証明終わり

誤りベクトルとハミング距離

符号語 $\mathbf{x} = (0110110), \mathbf{e} = (1000000)$ 1ビット誤りベクトル

$\mathbf{y} = \mathbf{x} + \mathbf{e} = (1110110)$

該当するビットが反転する

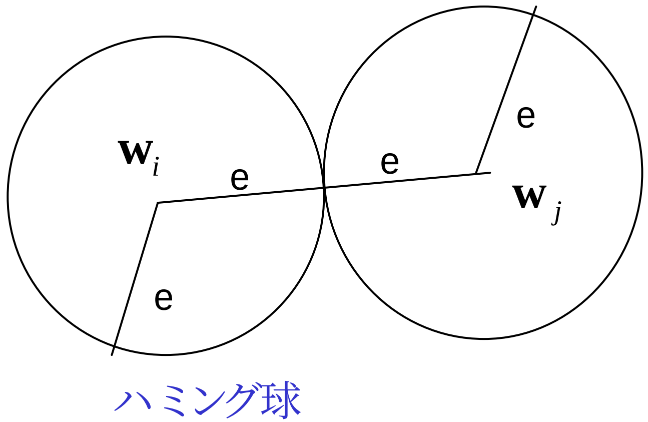
1 ビットのみ 1 であとはゼロのベクトル ‥‥**1 ビット誤りベクトル**
e ビットのみ 1 であとはゼロのベクトル ‥‥**e ビット誤りベクトル**

これらはハミング距離を用いて

$$d(\mathbf{x}, \mathbf{x} + \mathbf{e}) = 1 \quad \text{（eは1ビット誤りベクトル）}$$
$$d(\mathbf{x}, \mathbf{x} + \mathbf{e}) = e \quad \text{（eは}e\text{ビット誤りベクトル）}$$

ハミング球と誤り訂正可能条件

$$d_{\min} = \min_{\{\mathbf{w}_i, \mathbf{w}_j\}} d(\mathbf{w}_i, \mathbf{w}_j)$$

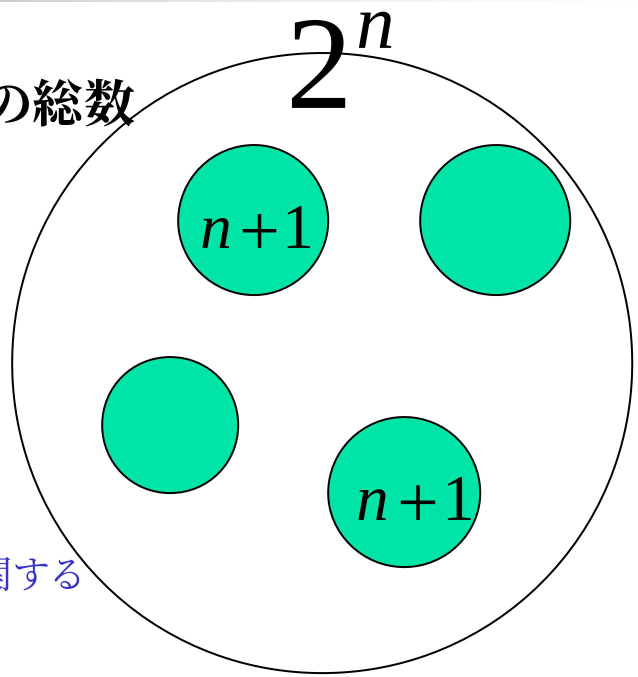


e ビット誤りが訂正できる条件

$$\boxed{d_{\min} \geq 2e + 1}$$

ハミングの不等式

1ビット誤りの生じたハミング球内にある系列の総数



$$1 + {}_nC_1 = 1 + n$$

オリジナルの符号語自身

1ビット誤りが訂正できるためには

$$\boxed{M \le \frac{2^n}{1+n}}$$

1ビット誤り訂正に関する
ハミングの不等式

これは容易に e ビット誤り訂正に拡張できて

$$\boxed{M \le \frac{2^n}{1+\sum_{i=1}^{e} {}_nC_i}}$$

e ビット誤り訂正に関する
ハミングの不等式