知能情報メディア概論II
情報量と暗号

システム情報系
岡本栄司

情報量

エントロピー $H(X)$　　確率事象

$$X=\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$: 情報源　a: 確率変数　p: 確率密度

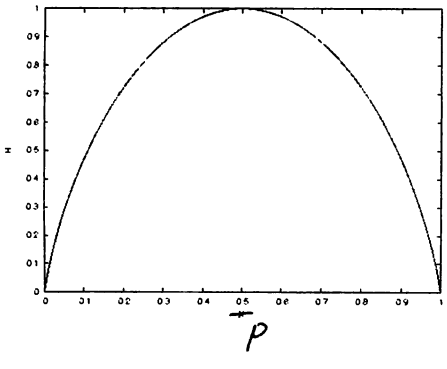
あいまいさ　$H(X)=-\sum_{i=1}^{n} p_i \log p_i$　(底が2のときbit)

例 n=2

$$H(X)=-p\log p-q\log q$$

$$p=P(X=a_1),\ q=P(X=a_2),\ p+q=1$$

$H(X)$



簡単な例

$n=2,\ p=q=1/2$のとき

$$H(X)=-\frac{1}{2}\log\frac{1}{2}-\frac{1}{2}\log\frac{1}{2}=1\ \text{(bit)}$$

$p_i=1/n$のとき

$$H(X)=-n\left(\frac{1}{n}\log\frac{1}{n}\right)=\log n\ \text{(bit)}$$

$n=2^k$なら$H(X)=k$ (bit)

等確率のとき、一番bit数が大きい。

ハミングコード
よくでてくるものに短いコードをわりあてる

各種エントロピー

$H(XY)=-\sum_{j=1}^{n}\sum_{i=1}^{n} p(a_i,b_j)\log p(a_i,b_j)$: 結合エントロピー

$$H(X|Y)=\sum_{j=1}^{m} p(b_j)\{-\sum_{i=1}^{n} p(a_i|b_j)\log p(a_i|b_j)\}$$

$$=-\sum_{j=1}^{n}\sum_{i=1}^{n} p(a_i,b_j)\log p(a_i|b_i)$$

$$=-\sum_{j=1}^{n}\sum_{i=1}^{n} p(a_i,b_j)\log\frac{p(a_i,b_j)}{p(b_j)}$$

$=H(XY)-H(Y)$: 条件付エントロピー

エントロピーの意味

$H(X)$: Xが持つ曖昧さ

$H(X|Y)=H(XY)-H(Y)$:

Yを知ったうえでなおXが持つ曖昧さ

$I(X;Y)=H(X)-H(X|Y)$:

Yを知ることによるXの曖昧さの減少分

=得た情報量: 相互情報量

C. E. Shannonの業績

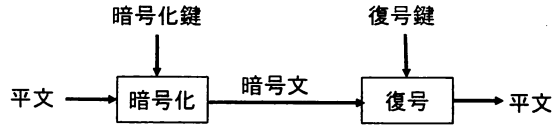
- 情報源符号化定理
  - 歪—なし
  - 歪—あり　（雑音）
- 通信路符号化定理　$C = W\log_2\left(1+\frac{S}{N}\right)$
- 暗号理論

レート
0.5
↓
0.4

エラーの割合を小さくするために(通信)レートを低くする必要ない？
4bit → 10bit で十分？

存在定理であって具体的にこうすればとは言ってない。

暗　号



暗号文
元のメッセージ → 完全暗号 ← 暗号文を入手したからって何も暗号文を入手していないのと同じ。

■ 定義　$H(M|C) = H(M)$　　鍵

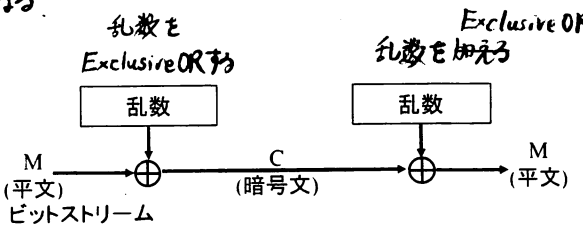
■ 定理　完全暗号ならば　$H(K) \geqq H(M)$　鍵が元のメッセージより長くなる。

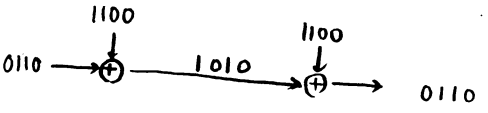
■ 利用公式 $H(XY) = H(X|Y) + H(Y)$
$= H(Y|X) + H(X)$

■ 証明

$H(M) = H(M|C) \leqq H(MK|C)$
$= H(M|KC) + H(K|C)$
$= H(K|C) \leqq H(K)$

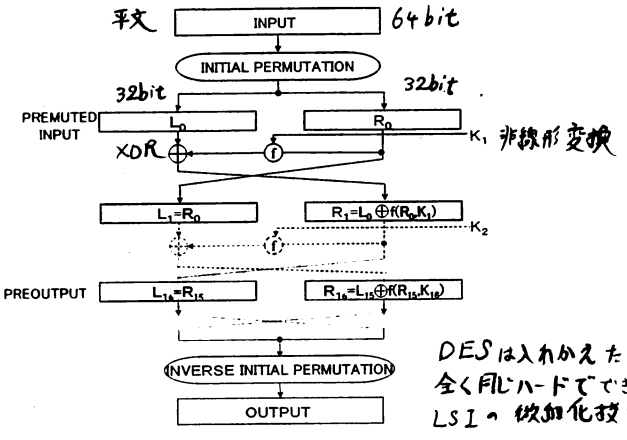
$H(M) \leq H(K)$ な暗号
完全暗号になっている。
CからM丼全くわからない。
バーナム暗号（どんなものも来る可能性がある）
$\exists C, \forall M : \exists K(C) = M$





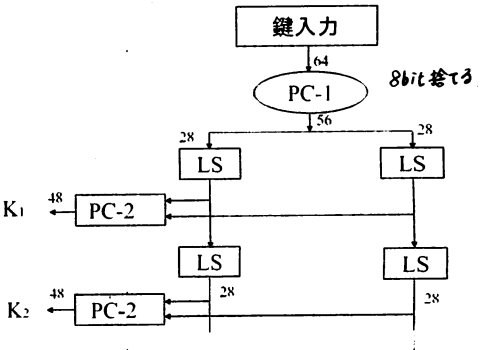
$0 \ll H(M|C) < H(M)$
十分大きければいいよね。

アメリカの標準暗号 DES (Data Encryption Standard)
1977年



DESは入れかえたりしているから全く同じハードでできる
LSIの微細化技術が…

鍵系列



公開鍵暗号系
RSA (Rivest, Shamir, Adleman) ← MITの人たち