北海道大学
Hokkaido University

情報エレクトロニクス学科共通科目・2年次・夏ターム〔必修科目〕

講義「情報理論」

第9回

第6章 通信路のモデル
第7章 通信路符号化の限界(1)

通信路符号化の基礎概念

通信路の統計的表現

- 通信路
  - 各時点において、一つの記号が入力され、一つの記号が出力される
  - 出力は入力から一意的に定まるのではなく、確率的に決まる

入力アルファベット $A=\{a_1,a_2,\ldots,a_r\}$ の元 $\xrightarrow{X_t}$ 通信路 Communication channel $\xrightarrow{Y_t}$ 出力アルファベット $B=\{b_1,b_2,\ldots,b_s\}$ の元

A=B かつ|A|=rのとき、r元通信路（r-ary channel）という

通信路の統計的性質は、任意の長さの<u>入力系列$X_0\ X_1\ \ldots,X_{n-1}$</u>とそれに対応する<u>出力系列$Y_0\ Y_1\ \ldots\ Y_{n-1}$ の確率分布</u>が与えられれば完全に定まる。

$$P_{Y_0Y_1\ldots Y_{n-1}|X_0\,X_1\ldots X_{n-1}}\,(y_0, y_1, \ldots, y_{n-1} \mid x_0, x_1, \ldots, x_{n-1})$$

=〔$X_0=x_0, X_1=x_1, \ldots, X_{n-1}=x_{n-1}$ であるとき、$Y_0=y_0, Y_1=y_1, \ldots, Y_{n-1}=y_{n-1}$ となる確率 〕

記憶のない定常通信路

記憶のない通信路(memoryless channel)**とは**

各時点の出力の現れ方が、その時点の入力には関係するが、それ以外の時点の入力・出力とは独立であるような通信路

$$P_{Y_0 \ldots Y_{n-1}|X_0 \ldots X_{n-1}}(y_0, \ldots, y_{n-1} \mid x_0, \ldots, x_{n-1}) = \prod_{i=0}^{n-1} P_{Y_i|X_i}(y_i \mid x_i)$$

記憶のない定常通信路とは

時間をずらしても統計的性質が変わらない記憶のない通信路

⇩

各時点において、入力Xが与えられたときの出力Yの条件付確率 $P_{Y|X}(y \mid x)$が同一

$$P_{Y_0 \ldots Y_{n-1}|X_0 \ldots X_{n-1}}(y_0, \ldots, y_{n-1} \mid x_0, \ldots, x_{n-1}) = \prod_{i=0}^{n-1} P_{Y|X}(y_i \mid x_i)$$

つまり、記憶のない定常通信路は$P_{Y|X}(y \mid x)$で決まる!

記憶のない定常通信路の表現法 ～通信路行列と通信路線図～

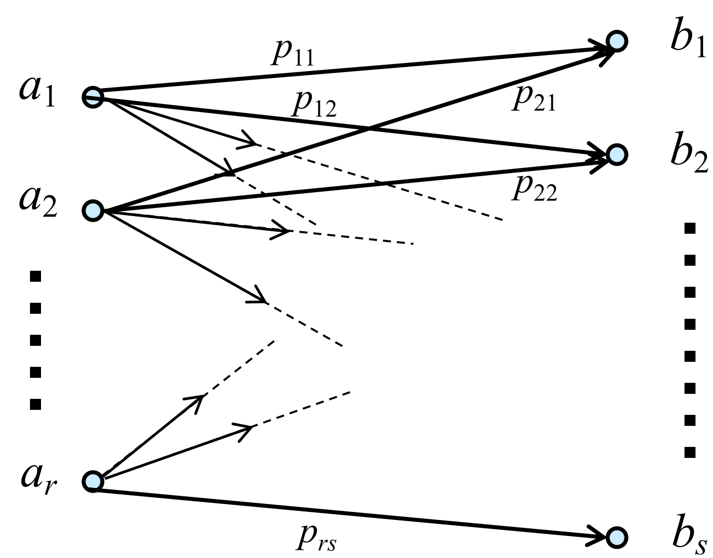
通信路行列とは

r 元入力アルファベット $A=\{a_1,a_2,\ldots,a_r\}$、
s 元出力アルファベット $B=\{b_1,b_2,\ldots,b_s\}$、
条件付確率 $p_{ij}=P_{Y|X}(b_j \mid a_i)$ としたとき

入力 $a_i \xrightarrow{p_{ij}} b_j$ 出力

条件付確率 p_{ij} を $(i\,,j\,)$ 要素とする $r\times s$ 行列

$$
T=\begin{array}{cc}
\begin{array}{cccc} b_1 & b_2 & \cdots & b_s \end{array} & \\
\begin{pmatrix}
p_{11} & p_{12} & \cdots & p_{1\,s} \\
p_{21} & p_{22} & \cdots & p_{2\,s} \\
\vdots & \vdots & & \vdots \\
p_{r\,1} & p_{r\,2} & \cdots & p_{r\,s}
\end{pmatrix} & \begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_r \end{array}
\end{array}
$$



図：通信路線図（channel diagram）

入出力に関して一様な通信路 ～記憶のない定常通信路の例～

入力に対して一様な通信路とは
通信路行列の各行が同じ要素の順序を入れ替えたものになっている通信路

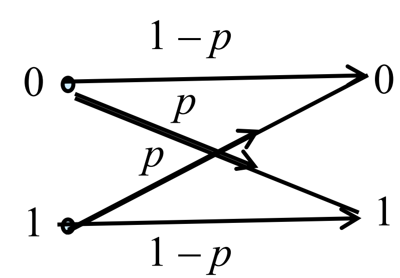
出力に関して一様な通信路とは
通信路行列の各列が同じ要素の順序を入れ替えたものになっている通信路

2重に一様な通信路とは　入力に対して一様でかつ出力に関して一様な通信路

- 例）2元対称通信路（binary symmetric channel; BSC）

2重に一様

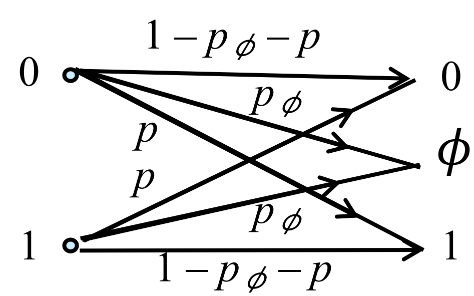
$$T = \begin{array}{cc} \begin{array}{cc} 0 & 1 \end{array} & \\ \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$



- 例）2元対称消失通信

入力に対して一様

$$T = \begin{array}{cc} \begin{array}{ccc} 0 & \phi & 1 \end{array} & \\ \begin{bmatrix} 1-p_\phi-p & p_\phi & p \\ p & p_\phi & 1-p_\phi-p \end{bmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}$$



ϕ : 消失（erasure）

2元通信路の誤りによる表現

- 入力アルファベット・出力アルファベットがともに $\{0,1\}$ の2元通信路は、誤り(error)を用いて表すことができる

誤り E を $\{0,1\}$ の元とすると、2元通信路の出力 Y は入力 X に誤り E を加えたものとみなせる！

$$Y = X \oplus E$$

- $E = 0$ 誤りなし
- $E = 1$ 誤り発生

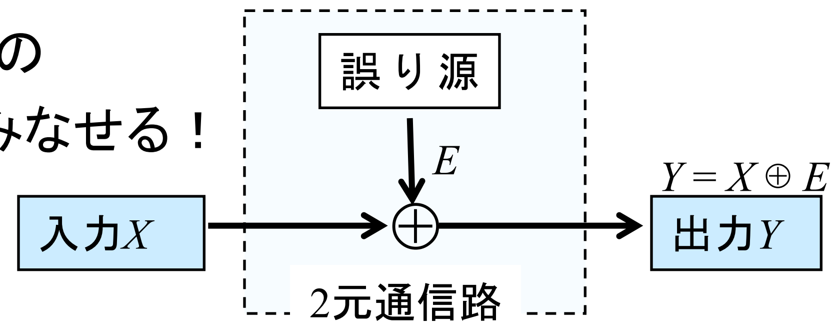
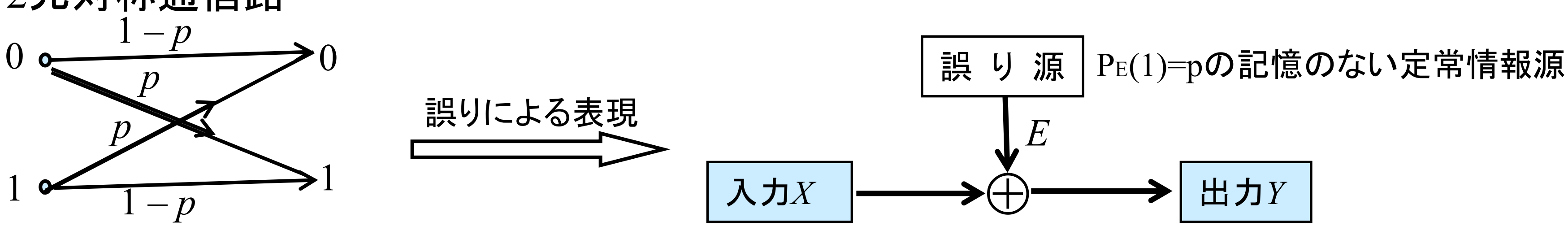


図: 誤りによる2元通信路の表現

※ふつう誤りの発生は入力と統計的に独立であると仮定される[加法的通信路]

例）2元対称通信路



このような誤りをランダム誤り(random error)という

誤りの発生確率 p をビット誤り率(bit error rate)と呼ぶ

バースト誤り通信路

バースト誤り通信路とは

誤りが一度生じると、その後しばらくの間は連続して誤りが発生すると考える**記憶のある通信路**（記憶のある誤り源で表される）の代表的なモデル

バースト誤り（burst error）とは

密集して生じる誤り

例えば、**誤り源から発信される系列**が $000\underbrace{11111111111}_{\text{バースト誤り}}000000000\underbrace{1111111}_{\text{バースト誤り}}0000\cdots$ という具合になる誤りのモデル

バースト誤り発生のマルコフモデルによる理解

- 誤り源が図のマルコフモデルで表せるような2元通信路を考える

$$\Pi = \begin{bmatrix} 1-\mathrm{P} & \mathrm{P} \\ p & 1-p \end{bmatrix}$$

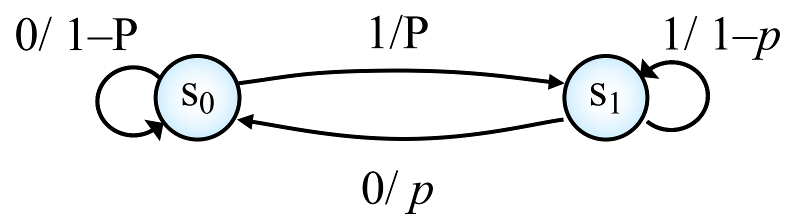


図: バースト誤り発生のマルコフモデル

定常分布を $\boldsymbol{w} = (w_0, w_1)$ とすると、
$\boldsymbol{w}\Pi = \boldsymbol{w}$ および $w_0+w_1 = 1$ から、

$$w_0 = \frac{p}{\mathrm{P}+p} \qquad w_1 = \frac{\mathrm{P}}{\mathrm{P}+p}$$

ビット誤り率

誤り源の出力 E が 1 となる確率 $P_E(1)$ を求めると･･･

$$P_E(1) = w_0\,\mathrm{P} + w_1\,(1-p) = \frac{1}{\mathrm{P}+p}\{p\,\mathrm{P} + \mathrm{P}\,(1-p)\} = \frac{\mathrm{P}}{\mathrm{P}+p} = w_1$$

バースト誤りの長さの平均(1/2)

- P.8の図の2元通信路で発生したバースト誤りの平均長は？
  - 誤り系列における 1 の連続（1のラン）を任意に一つ取り出す
  - その長さがℓ となる確率 $P_{\mathrm{B}}(\ell)$ を求めると、

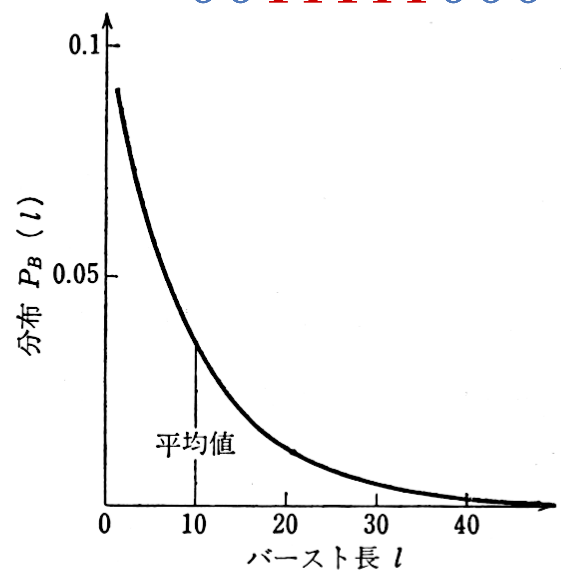
$$P_{\mathrm{B}}(\ell) = \underbrace{(1-p)^{\ell-1}}_{\text{最初の1の後に、1が}\ell-1\text{連続する確率}}\ \underbrace{p}_{\text{0が出る確率}} \quad (\ell = 1, 2, \cdots)$$

となる

  - バースト誤りの長さ（バースト長）の平均値$\overline{\ell}$ は次のようになる

$$\overline{\ell} = \sum_{\ell=1}^{\infty} \ell\, P_{\mathrm{B}}(\ell) = \sum_{\ell=1}^{\infty} \ell\,(1-p)^{\ell-1} p = \frac{1}{p}$$

・・・0011111000・・・



図：バースト長の分布の例
(P=0.0001,p=0.1)

バースト誤りの長さの平均(2/2)

$\overline{\ell}'$:同じビット誤率 $P/(P+p)$ のランダム誤り通信路のバースト誤りの長さの平均長

$\overline{\ell}' < \overline{\ell}$

実際、$p'=P/(P+p)$*とすれば、*

$$\overline{\ell}' = \sum_{\ell=1}^{\infty} \ell\,(p')^{\ell-1}(1-p') = 1/(1-p') = (P+p)/p$$

となる。例えば$P=0.001$, $p=0.1$ *であれば、*

$$\overline{\ell} = 1/p = 1/0.1 = 10$$
$$\overline{\ell}' = (P+p)/p = 0.101/0.1 = 1.01$$

となり、$\overline{\ell}' < \overline{\ell}$*が成り立っている。*

通信が不安定な状態を表す誤りモデル

- **ギルバートモデル**（Gilbert model）
  - p.8の図のモデルでは、バースト誤りの発生期間はすべての記号が誤る（**ソリッドバースト誤り**（solid burst error））
  - 正誤が混在するバースト誤りの方が自然
  - 状態Bのときは
    1,0 をそれぞれ h, $1-h$
    の確率で発生させる
    （状態遷移と記号出力を分離）

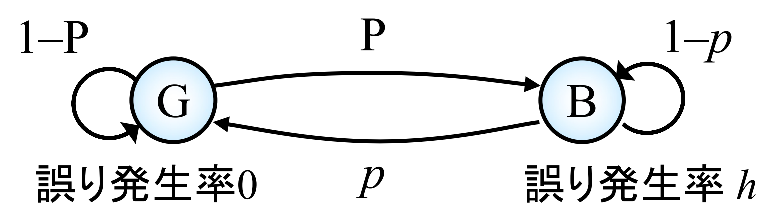


図 ギルバートモデル

ビット誤り率

定常分布において状態G、Bにいる確率をw_G、w_Bとすれば

$$P_E(1) = w_G \times 0 + w_B \times h = \frac{p}{P+p} \times 0 + \frac{P}{P+p} \times h = \frac{Ph}{P+p}$$

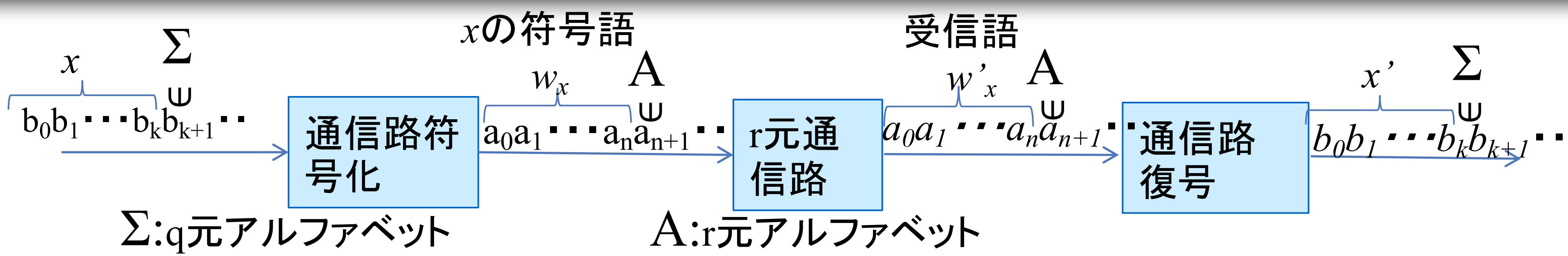
p.8の図のモデルより小さい

バースト誤り（状態Bに連続的に留まる現象）の長さの平均値 $\overline{\ell}$ は

$$\overline{\ell} = 1/p$$

p.8の図のモデルと同じ

通信路符号の基礎概念(1)



- 通信路符号化の目的: 信頼性の向上　そのために→冗長性を付加
- 長さkのブロック$x\in\Sigma^k$を長さnの符号語$w_x\in A^n$に符号化
- $q^k < r^n$: 符号語としてA^nの一部のみ使用 (冗長性の付加)

[例]$\Sigma=\{0,1\}$, $k=1$,$A=\{0,1\}$,$n=3$ とすれば、

$$A^3=\{000, 001, 010, 011, 100, 101, 110, 111\}$$

この内、000 と 111 の二つだけを符号語として用いる

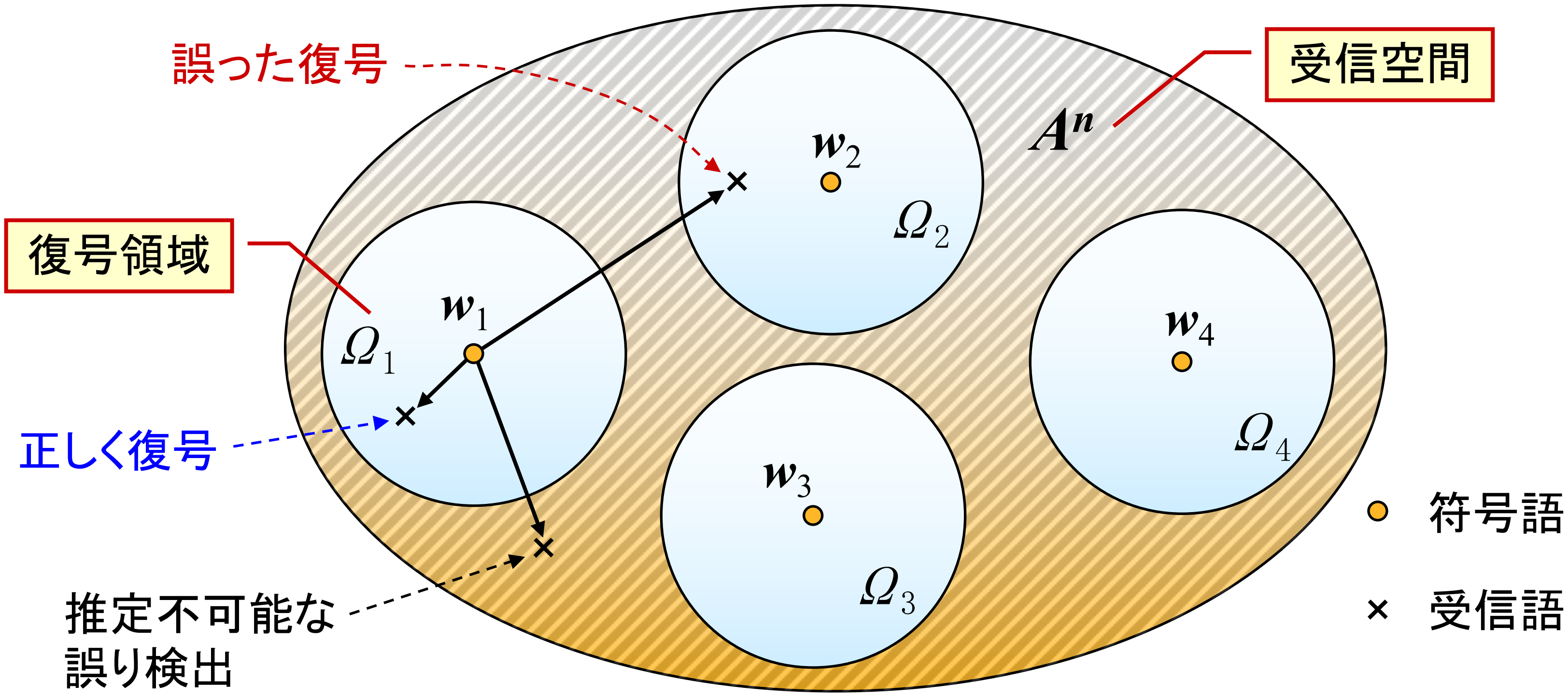
$$0 \rightarrow 000, \qquad 1 \rightarrow 111$$

このとき、送った符号語w_xに1ビット誤りが生じた受信語w'_xを受け取っても、w'_x内の3ビットの多数決により訂正可能

$$001, 010, 100 \rightarrow 0, \qquad 110, 101, 011 \rightarrow 1$$

通信路符号の基礎概念(2)

- **通信路符号**(または**符号**): $\mathbf{A}^n$ の中から選ばれた系列の集合 C
- 図は通信路符号による誤り訂正の概念図



図(通信路符号の)復号の概念図

情報速度と冗長度

- 符号C(⊆A^n)に含まれる符号語の数をM(=q^k)とする。
  符号語を等確率で用いると、符号語の平均情報量は$\log_2 M$ビットとなるから
  $$R = \frac{\log_2 M}{n}$$ （ビット／記号）
  の速度で情報を伝達できることになる。この R をこの符号の情報伝送速度または単に情報速度と呼ぶ。情報速度Rを用いれば、符号語数 M は
  $$M = 2^{nR}$$
- A^n に含まれる r^n 個の系列すべてを符号語とすれば、情報速度は最大値
  $$R_{\mathrm{max}} = (\log_2 r^n) / n = \log_2 r$$
  となる。$R < R_{\mathrm{max}}$ とすることで、誤りの訂正や検出が可能となる。
- 情報速度 R の符号 C に対し、
  $$\eta = R / R_{\mathrm{max}}$$
  を、C の効率または符号化率(code rate)と呼ぶ。
  これは、$0 < \eta < 1$ を満たす。また、
  $$\rho = 1 - \eta$$
  を C の冗長度という。

すべての系列を使うと
誤りを検知できない

冗長度≒信頼性
冗長度と効率はTrade-off

最尤復号法(1)

- 符号語 $w_1, w_2, \ldots, w_M$ に対する復号領域 $\Omega_1, \Omega_2, \ldots, \Omega_M$ の定め方は？
- P_C: 正しく復号される確率

  最尤復号法=

  (符号語が等確率で送られるとき)P_C を最大にするような復号領域の定め方

- 符号語 w_i を送ったとき
  正しく復号される確率は

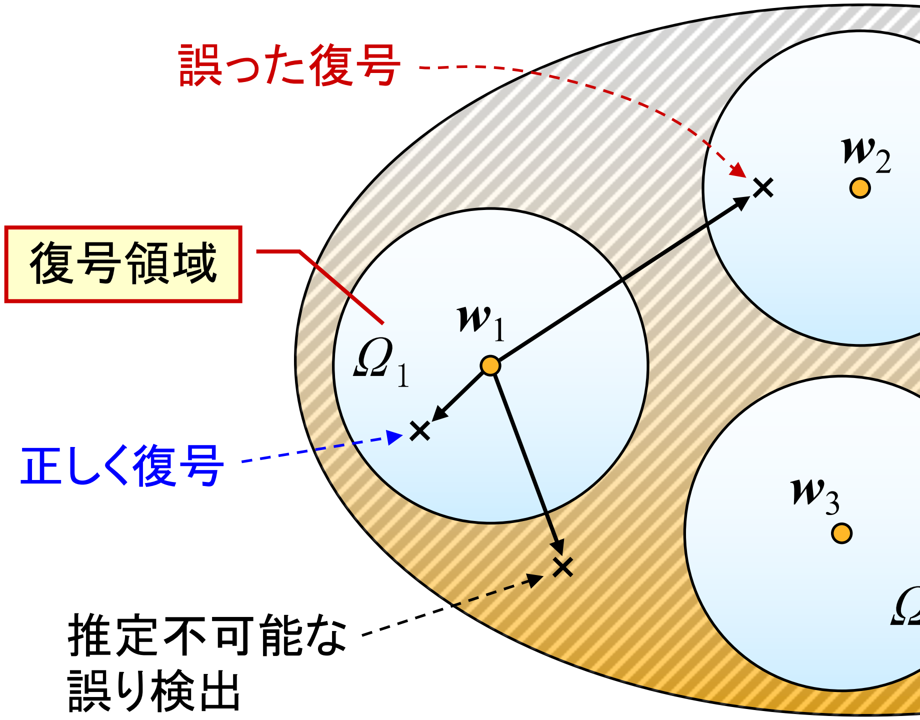
$$P_C(w_i) = \sum_{y \in \Omega_i} P(y \mid w_i)$$

となる。ただし

$P(y \mid w_i)$: w_i を送ったとき

y が受信される確率

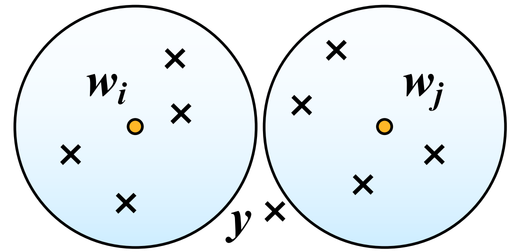
とする。



最尤復号法(2)

- どの符号語が送られてくる確率も等しく $1/M$ であると仮定すると
  正しく復号される確率P_Cは

$$P_C=\sum_{i=1}^{M}\frac{1}{M}\times P_C(w_i)=\frac{1}{M}\sum_{i=1}^{M}\sum_{y\in\Omega_i}P(y\mid w_i)$$



と書ける。これを最大にするには、

$$\Omega_i=\{y\in A^n \mid P(y\mid w_i)\geqq P(y\mid w_j)\ \text{for all}\ j=1,\ldots,M\}$$

とすればよい。実際、Ω_i に属するある y に対し、

$$P(y\mid w_i)<P(y\mid w_j)$$

となったとすれば、この y を Ω_j に移すことで、$P(y\mid w_j)-P(y\mid w_i)>0$ だけP_C が増加。

- y を受け取ってw_i を推定する問題とすれば、$P(y\mid w_i)$ はw_iの尤度である。
  尤度を最大化する復号法を最尤復号法(maximum likelihood decoding)と呼ぶ。

- 最尤復号法は、各符号語が等確率で送られるとき、正しく復号される
  確率 P_C を最大とするという意味で、最良の復号法である。しかし、すべての
  符号語 w_i に対し、$P(y\mid w_i)$ を計算し比較し(orその結果を表にもた)なければならず、
  符号語数 M が大きい場合には、実用的ではない。