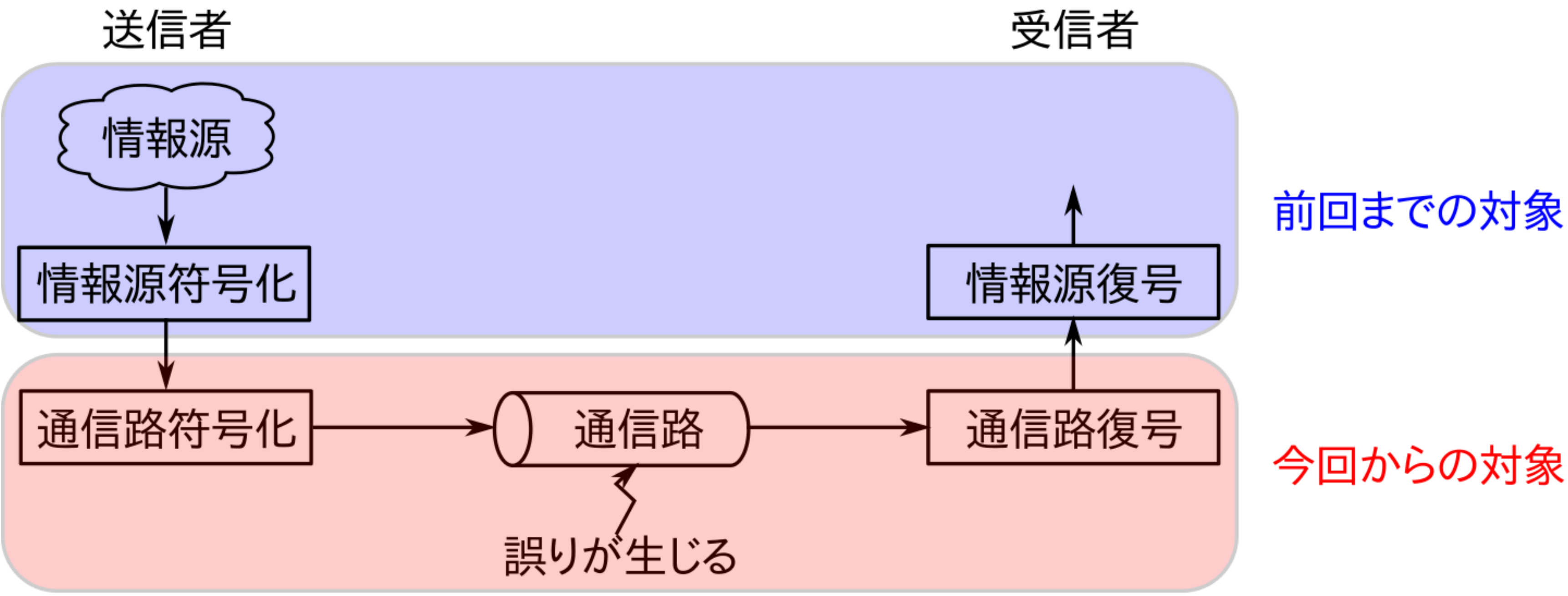
情報理論 第 6 回
– 通信路モデル・通信路容量・通信路符号化の基礎 –
(教科書: 11 章 , 14.2 節, 14.3 節)

野崎 隆之

0. 導入 (1)

今回 (第 6 回) から第 7 回までで通信路符号化について学ぶ

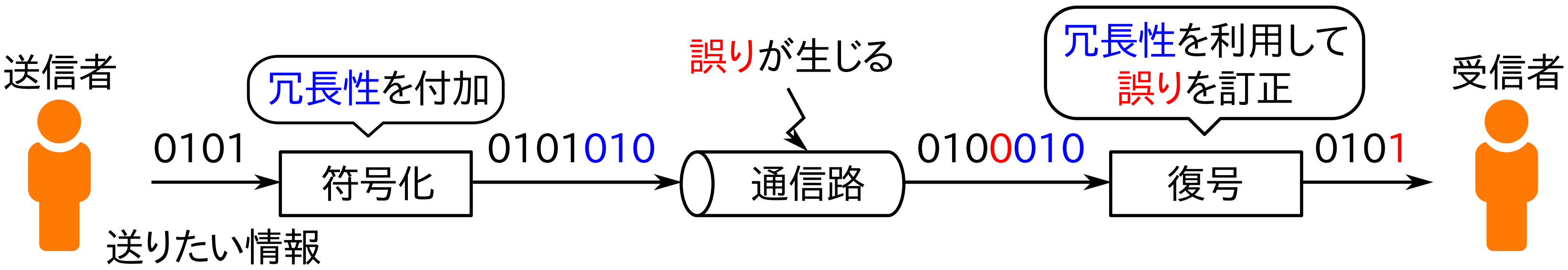


今後の流れ

- 通信路を確率的にモデル化・モデルの数学的な性質 (第 6 回)
- 符号化によってどのぐらいの効率で情報を送れるか?
(通信路容量・情報源符号化定理・第 6 回)
- 誤りを減らすための符号化法
(通信路符号化・誤り訂正符号・符号理論・第 7 回)

0. 導入 (2)

通信路符号化の目的 (復習)



- 誤りの生じる通信路を使って誤りのない通信をしたい
  ⇒ 復号誤り率 (誤った情報が伝わる確率) を 0 に近づけたい
- なるべく効率の良い通信を行いたい
  ⇒ 冗長性をなるべく減らしたい
- 復号誤り率 0 のときにどこまで冗長性を下げられるか?
  (通信路符号化定理)

0. 導入 (3)

今日の目標

- 通信路を確率的にモデル化できる
- 通信路容量について計算・説明ができる
- 通信路符号化・通信路符号化定理について説明できる

今日の流れ

1. 通信路モデル
   (条件付き確率を用いて記述)
2. 通信路容量
   (1 回の通信で得られる相互情報量の最大値 $\approx$ 伝送効率の限界)
3. 通信路符号化の基礎

1. 通信路モデル (1: 概要)

通信路：通信・記録における情報の通り道 (たまに誤りが生じる)

通信路モデル

通信路で生じる誤りを確率的に記述したもの

- 情報の伝達では媒体の物理的性質を利用
  例：無線通信 — 電磁波の波長・振幅・位相，ハードディスク — 磁界の向き
- 誤りの生じ方は情報伝達をする媒体に依存

この話題の流れ

1. 通信路の例
2. 通信路を数学的に記述
3. いくつかの通信路モデルの例
   - 二元対称通信路
   - 二元消失通信路
   - Z 通信路

1. 通信路モデル (2: 通信路の例 1)

(例) より対線

- LAN ケーブル，電話線 etc で利用
- 銅線中の電気信号で 0,1 を表現 (0: 正の電圧，1: 負の電圧)
- 周囲の電磁波 (e.g. 電源ケーブル) の影響で雑音が発生

(例) 無線伝送

- 携帯電話，無線 LAN，テレビ放送 etc で利用
- 電磁波の振幅・周波数・位相で 0,1 を表現 (e.g. 0: 振幅小，1: 振幅大)
- 周囲の電磁波の影響で雑音が発生

より詳しくは「情報ネットワーク」(3 年生前期) で！

1. 通信路モデル (3: 通信路の例 2)

(例) 磁気記録

- ハードディスク，磁気テープ
- 磁性体の向き (磁界の向き) で 0,1 を表現
- 周囲の磁性体の影響で向きが反転

(例) フラッシュメモリ (不揮発性メモリ)

- Solid-State Disc (SSD), USB メモリ
- 電界効果トランジスタ中の電荷量で 0,1 を表現 (0: 電荷なし， 1: 電荷あり)
- 電荷の増減によって誤りが生じる

1. 通信路モデル (4: 定義 1)

(定義) 入力アルファベット，出力アルファベット

- $\mathcal{X}$: 入力アルファベット
  通信路の入力シンボルの集合 (この講義では $\mathcal{X} = \{0, 1\}$)
- $\mathcal{Y}$: 出力アルファベット
  通信路の出力シンボルの集合

一般の通信路モデル

- $x_1 x_2 \cdots x_k \in \mathcal{X}^k$: 入力系列
- $y_1 y_2 \cdots y_k \in \mathcal{Y}^k$: 出力系列
- x_t, y_t : 時刻 t の入力，出力

$$P_{Y_1 Y_2 \cdots Y_k | X_1 X_2 \cdots X_k}(y_1 y_2 \cdots y_k | x_1 x_2 \cdots x_k)$$

入力系列が与えられた時に，出力系列の確率分布を与えている.

情報源モデルと同様に一般性の高すぎるモデルは扱いづらい...

1. 通信路モデル (4: 定義 2)

定常無記憶通信路

- (無記憶性) 時刻 t の出力 y_t は時刻 t の入力 x_t のみに依存
- (定常性) 通信路の統計的性質が時刻 t に依存しない

$$\begin{aligned} & P_{Y_1 Y_2 \cdots Y_k | X_1 X_2 \cdots X_k}(y_1 y_2 \cdots y_k | x_1 x_2 \cdots x_k) \\ &= \prod_{t=1}^{k} P_{Y_t | X_t}(y_t | x_t) \qquad \text{(無記憶性)} \\ &= \prod_{t=1}^{k} P_{Y|X}(y_t | x_t) \qquad \text{(定常性)} \end{aligned}$$

定常無記憶通信路の場合, $P_{Y|X}(y|x)$ のみを記述すれば良い

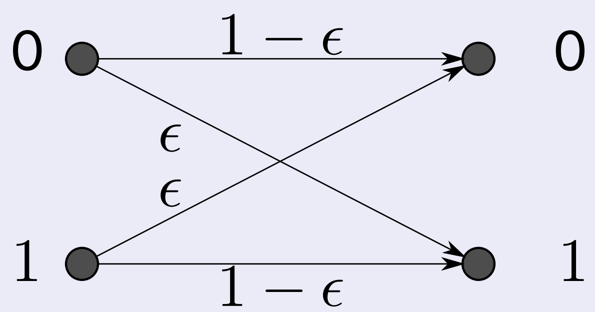
1. 通信路モデル (6: 二元対称通信路)

二元対称通信路 (BSC: Binary Symmetric Channel)

- $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1\}$
- 正しく情報が送られる確率が $(1 - \epsilon)$，誤った情報が届く確率 ϵ

$$
\begin{aligned}
P_{Y|X}(0|0) &= 1 - \epsilon, & P_{Y|X}(1|0) &= \epsilon, \\
P_{Y|X}(0|1) &= \epsilon, & P_{Y|X}(1|1) &= 1 - \epsilon,
\end{aligned}
$$

- ϵ を反転確率と呼ぶ $(0 \leq \epsilon \leq 1/2)$



BSC の通信路線図

最も基本的な誤りのある通信路

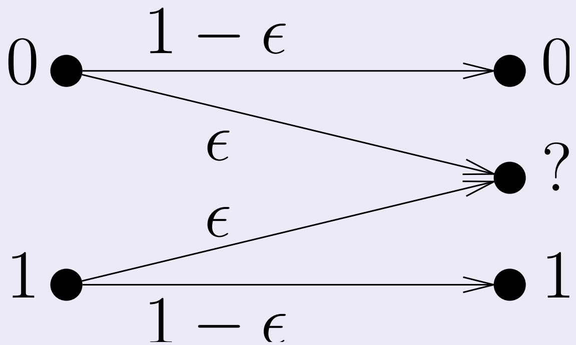
1. 通信路モデル (7: 二元消失通信路)

二元消失通信路 (BEC: Binary Erasure Channel)

- $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, ?\}$
- 正しく情報が送られる確率が $(1 - \epsilon)$，情報が消失する確率 ϵ

$$
\begin{aligned}
&P_{Y|X}(0|0) = 1 - \epsilon, \quad && P_{Y|X}(?|0) = \epsilon \quad && P_{Y|X}(1|0) = 0, \\
&P_{Y|X}(0|1) = 0, && P_{Y|X}(?|1) = \epsilon && P_{Y|X}(1|1) = 1 - \epsilon,
\end{aligned}
$$

- ϵ を消失確率と呼ぶ $(0 \leq \epsilon \leq 1)$



BEC の通信路線図

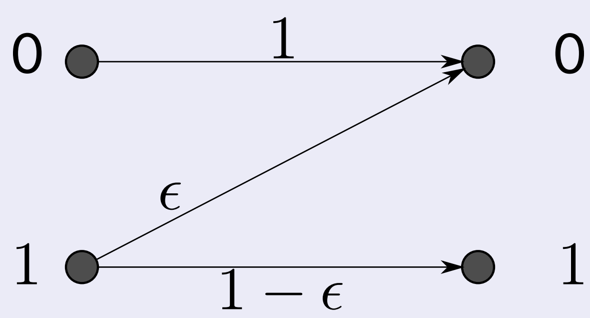
最も簡単な通信路モデル

1. 通信路モデル (8: Z 通信路)

Z 通信路

- $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, \}$
- 0 は必ず正しく送られるが，1 は確率 ϵ で 0 に遷移
- ϵ を (通信路の) 誤り率と呼ぶ

$$
\begin{aligned}
P_{Y|X}(0|0) &= 1, \quad & P_{Y|X}(1|0) &= 0, \\
P_{Y|X}(0|1) &= \epsilon, \quad & P_{Y|X}(1|1) &= 1 - \epsilon,
\end{aligned}
$$



Z 通信路の通信路線図

最も単純な非対称通信路

2. 通信路容量 (1: 概要)

通信路が与えられた時に相互情報量 $I(X;Y)$ をどこまで大きく出来るか?

- 通信路 $P_{Y|X}(y|x)$ はシステムで決まる (設計できない)
- 入力分布 $P_X(x)$ は送信者が決められる (設計可能)

通信路容量：

- 入力分布 $P_X(x)$ を変化させた時の $I(X;Y)$ の最大値
- 与えられた通信路の伝送速度の上限 (後で詳しく)

(復習) 通信路モデル (定常無記憶通信路)

- X : 入力アルファベットに関する確率変数
- Y : 出力アルファベットに関する確率変数
- $P_{Y|X}(y|x)$: 定常無記憶通信路

(復習) 相互情報量

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

2. 通信路容量 (2: 定義)

通信路容量 (Channel Capacity)

与えられた通信路 $p_{Y|X}(y|x)$ に対して，通信路容量 C は

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

通信路容量の計算手順

1. 入力分布 $P_X(x)$ を変数 $\boldsymbol{q}$ を用いて表記
   二元入力 ($X \in \{0,1\}$) の場合，$P_X(0) = q, P_X(1) = 1-q$
   [k 元入力の場合は多変数で表記 $P_X(i) = q_i \quad (\sum_i q_i = 1)$]
2. 入力分布 $P_X(x)$ と通信路 $P_{Y|X}(y|x)$ から $I(X;Y)$ を求める
3. 変数 $\boldsymbol{q}$ について $I(X;Y)$ を最大化

2. 通信路容量 (3: 例)

(例) 二元対称通信路の通信路容量

1. 入力分布 $P_X(x)$ を変数 $\boldsymbol{q}$ を用いて表記
   二元入力 ($X \in \{0,1\}$) の場合，$P_X(0) = q, P_X(1) = 1 - q$
2. 入力分布 $P_X(x)$ と通信路 $P_{Y|X}(y|x)$ から $I(X;Y)$ を求める
   $h_2(x) := -x \log x - (1-x)\log(1-x)$

$$I(X;Y) = h_2(q + \epsilon - 2q\epsilon) - h_2(\epsilon)$$

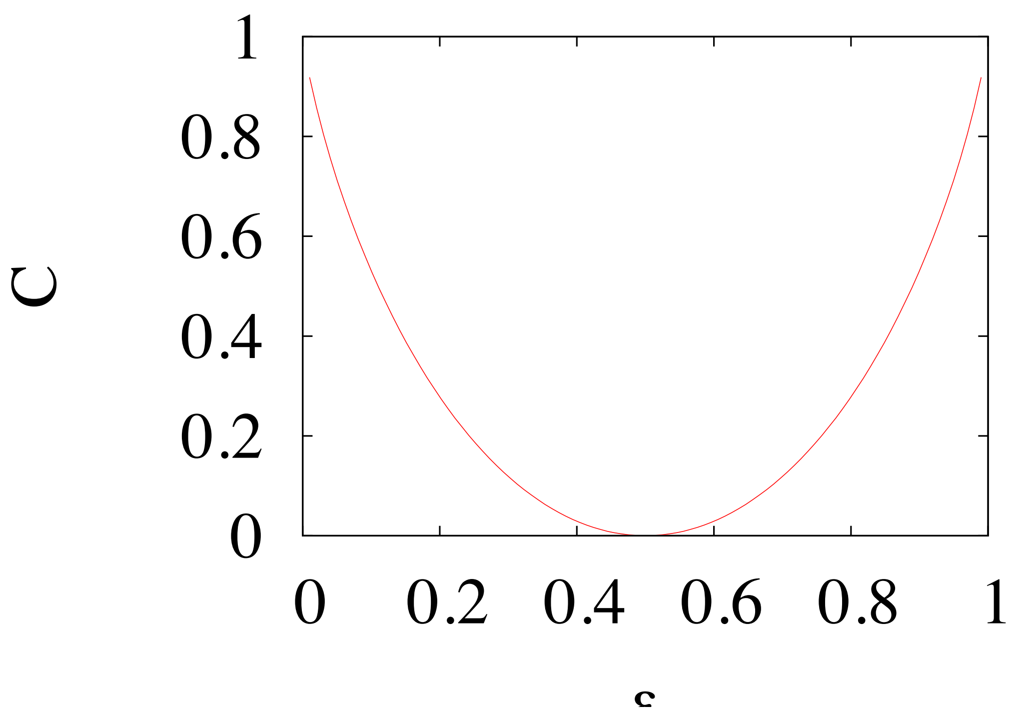
3. 変数 $\boldsymbol{q}$ について $I(X;Y)$ を最大化 (第 1 回レポートの結果を利用)

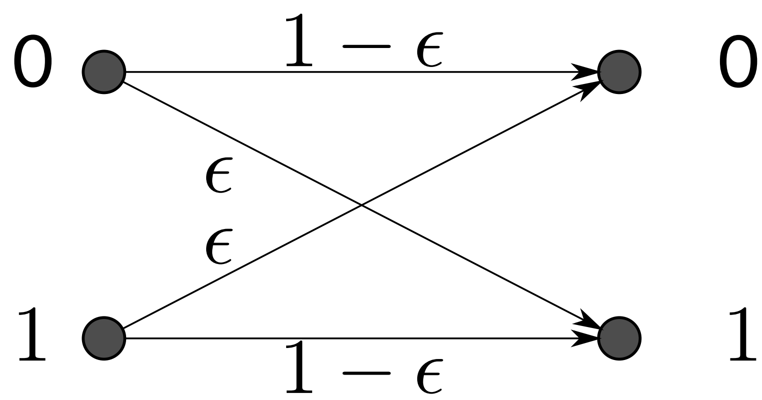
$$\begin{aligned} C &= \max_q [h_2(q + \epsilon - 2q\epsilon) - h_2(\epsilon)] \\ &= \max_q [h_2(q(1-2\epsilon) + \epsilon)] - h_2(\epsilon) \\ &= 1 - h_2(\epsilon) \quad (h_2(x) \text{ は } x = 1/2 \text{ のとき最大値 } 1 \text{ をとる}) \end{aligned}$$

(最大値は $q = 1/2$ の時)

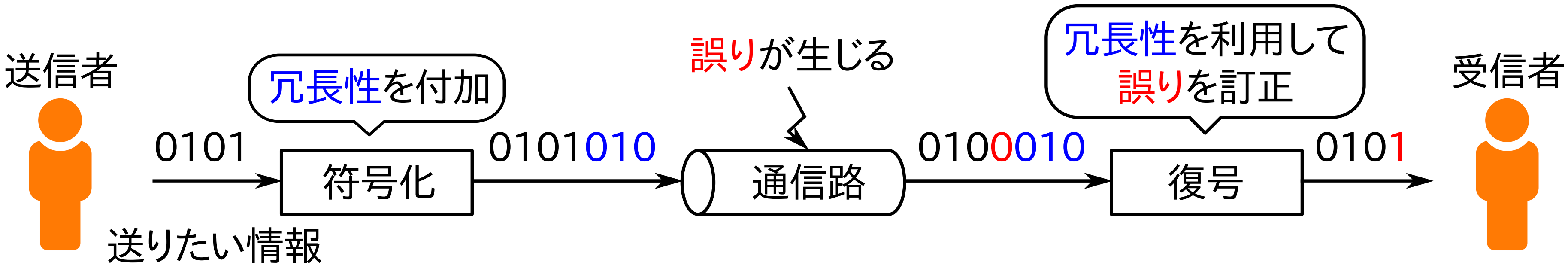
2. 通信路容量 (4: 結果の考察)

- $q = 1/2$ の時に最大値
  0,1 を等確率に入力すると通信路を最大限利用できる
- $\epsilon = 1/2$ のとき $C = 0$
  出力 Y から入力 X の情報を一切得られない
- $\epsilon = 0, 1$ のとき $C = 1$
  誤りなく通信が可能．($\epsilon = 1$ の時は出力結果を反転させる)





3. 通信路符号化 (1: 概要)



モチベーション

- 通信路で生じる誤りを訂正したい (誤り訂正符号)
- 通信路で生じる誤りを検出したい (誤り検出符号)
- 通信路で生じる消失を訂正したい (消失訂正符号)

方法

送信情報に冗長性を付与する

3. 通信路符号化 (2: 通信路符号)

(定義) 通信路符号

- X : 通信路の入力アルファベット
- X^n : 長さ n の入力アルファベットの系列
- $\mathcal{C} \subset X^n$: 通信路符号
- n : 符号長
- $|\mathcal{C}|$: 符号語数 ($|C|$ は C の要素数)

$\mathcal{C}$ の要素を符号語と呼ぶ

(例) $n = 3$ の繰り返し符号

- $X := \{0, 1\}$, $X^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$
- $\mathcal{C} = \{000, 111\}$
- 符号長 3
- 符号語数 2

3. 通信路符号化 (3: 符号化・復号)

符号化

メッセージ $\boldsymbol{m} \in \mathcal{M}$ を符号語 $\boldsymbol{x} \in \mathcal{C}$ に割り当てる操作を符号化と呼ぶ

$$\phi : \mathcal{M} \to \mathcal{C}$$

特に，$\phi(\boldsymbol{m})$ を $\boldsymbol{m}$ の符号語と呼ぶ

注意： $|\mathcal{M}| = |\mathcal{C}|$

復号

受信語 $\boldsymbol{y} \in Y^n$ からメッセージ $\boldsymbol{m} \in \mathcal{M}$ を推定する操作を復号と呼ぶ

$$\psi : \mathcal{Y}^n \to \mathcal{M}$$

誤り検出符号の場合は，$\psi : \mathcal{Y}^n \to \mathcal{M} \cup \perp$

注意：別のメッセージに復号してしまうことを誤訂正と呼ぶ

3. 通信路符号化 (4: 符号化・復号の例)

(例) $n = 3$ の繰り返し符号

符号化

$$\phi(0) = 000, \qquad \phi(1) = 111$$

復号（多数決復号法）

$$\psi(000) = \psi(001) = \psi(010) = \psi(100) = 0,$$
$$\psi(111) = \psi(110) = \psi(101) = \psi(011) = 1$$

0 に復号される受信後の集合を 0 の復号領域と呼び，$\Omega(0)$ と書く

性能の良い誤り訂正符号を作るには

- 符号 $\mathcal{C}$
- 復号法 (復号領域)

をうまく設計する必要がある.

3. 通信路符号化 (5: 性能指標 1)

(定義) 伝送速度（符号化率，レート）

$$R = \frac{1}{n} \log_{|X|} |\mathcal{C}|$$

(直感的には，符号長のうちメッセージに相当する部分の割合)

(例) $n = 3$ の繰り返し符号

$$R = \frac{1}{3} \log_2 2 = \frac{1}{3}$$

(定義) 復号誤り率

受信語から正しくメッセージが推定できない確率 P_E

(例) $n = 3$ の繰り返し符号 + 多数決復号

2 つ以上間違えると正しく復号できないので，$P_E = 3\epsilon^2(1-\epsilon) + \epsilon^3$

3. 通信路符号化 (6: 性能指標 2)

符号語同士の近さを測りたい
(遠ければ遠いほど誤り訂正能力が高い)

(定義) ハミング距離

ベクトル $(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n)$ に対して,
ハミング距離 $d_H(\boldsymbol{a}, \boldsymbol{b})$ は

$$d_H(\boldsymbol{a}, \boldsymbol{b}) := \sum_{i=1}^{n} \mathbb{I}[a_i \neq b_i] \qquad \mathbb{I}[Q] = \begin{cases} 1 & \text{命題 } Q \text{ が真} \\ 0 & \text{命題 } Q \text{ が偽} \end{cases}$$

(直感的には, 食い違っているシンボルの個数)

$$d_H(00100, 10110) = 2$$

3. 通信路符号化 (7: 性能指標 3)

符号語の組の中で最も距離が近いもの (復号性能の指標になる)

(定義) 最小距離

符号 $\mathcal{C}$ に対して，最小距離 $d_{\min}(\mathcal{C})$ は

$$d_{\min}(\mathcal{C}) := \min\{d_H(\boldsymbol{u}, \boldsymbol{v}) \mid \boldsymbol{u}, \boldsymbol{v} \in \mathcal{C}\}$$

$\mathcal{C} = \{0000, 0011, 1110, 1111\}$
$d_{\min}(\mathcal{C}) = 1$

3. 通信路符号化 (7: 通信路符号の例題)

(例題)

$\mathcal{C} = \{00000, 01110, 10111, 11000\}$

- 符号長は?
- 符号語数は?
- 伝送速度は?
- 最小距離は?

3. 通信路符号化 (8: 繰り返し符号の限界)

$n = 2t+1$ の繰り返し符号

- $\mathcal{C} = \{00\cdots 0, 11\cdots 1\}$
- 復号法：多数決復号 (t 個まで訂正可能)
- 伝送速度 $R^{(t)} = 1/t$
- 復号誤り率 $P_E^{(t)} = \sum_{i=t+1}^{2t+1} \binom{2t+1}{i} \epsilon^i (1-\epsilon)^{2t+1-i}$
  $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ (2 項係数)

$\lim_{t\to\infty} P_E^{(t)} = 0$ であるが，$\lim_{t\to\infty} R^{(t)} = 0$
$\Rightarrow$ 誤り率を 0 にできるが，伝送速度も 0 になってしまう...

4. 通信路符号化定理

(定理 1) 通信路符号化順定理

C: 通信路容量
$R < C$ のとき，$P_E \to 0$ $(n \to \infty)$ となる通信路符号が存在する

(定理 2) 通信路符号化 (強) 逆定理

$R > C$ のとき，どんな通信路符号と復号の組を与えても，
$P_E \to 1$ $(n \to \infty)$ となる.

意味：どんなに良い通信路符号を作っても，伝送速度は通信路容量までしか到達しない
証明の方法： 色々な手法があるが時間がかかるので割愛

- 植松「情報理論の考え方」(pp.159–180, 典型系列を利用)
- 坂庭・笠井「通信理論入門」(pp.124–142, ランダム符号を利用)

今日のまとめ (1)

定常無記憶通信路

- (無記憶性) 時刻 t の出力 y_t は時刻 t の入力 x_t のみに依存
- (定常性) 通信路の統計的性質が時刻 t に依存しない

$$
\begin{aligned}
& P_{Y_1Y_2\cdots Y_k|X_1X_2\cdots X_k}(y_1y_2\cdots y_k|x_1x_2\cdots x_k) \\
&= \prod_{t=1}^{k} P_{Y_t|X_t}(y_t|x_t) \qquad \text{(無記憶性)} \\
&= \prod_{t=1}^{k} P_{Y|X}(y_t|x_t) \qquad \text{(定常性)}
\end{aligned}
$$

- 二元対称通信路
- 二元消失通信路
- Z 通信路

今日のまとめ (2)

通信路容量 (Channel Capacity)

与えられた通信路 $p_{Y|X}(y|x)$ に対して，通信路容量 C は

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

通信路符号化

- 符号化 $\phi : \mathcal{M} \to \mathcal{C}$
- 復号 $\psi : \mathcal{Y}^n \to \mathcal{M}$
- 伝送速度 $R = \frac{1}{n} \log_{|X|} |\mathcal{C}|$
- 復号誤り率： 受信語から正しくメッセージが推定できない確率 P_E
- 最小距離 $d_{\min}(\mathcal{C}) := \min\{d_H(\boldsymbol{u}, \boldsymbol{v}) \mid \boldsymbol{u}, \boldsymbol{v} \in \mathcal{C}\}$

今日のまとめ (3)

(定理 1) 通信路符号化順定理

C: 通信路容量
$R < C$ のとき，$P_E \to 0$ $(n \to \infty)$ となる通信路符号が存在する

(定理 2) 通信路符号化 (強) 逆定理

$R > C$ のとき，どんな通信路符号と復号の組を与えても，
$P_E \to 1$ $(n \to \infty)$ となる.