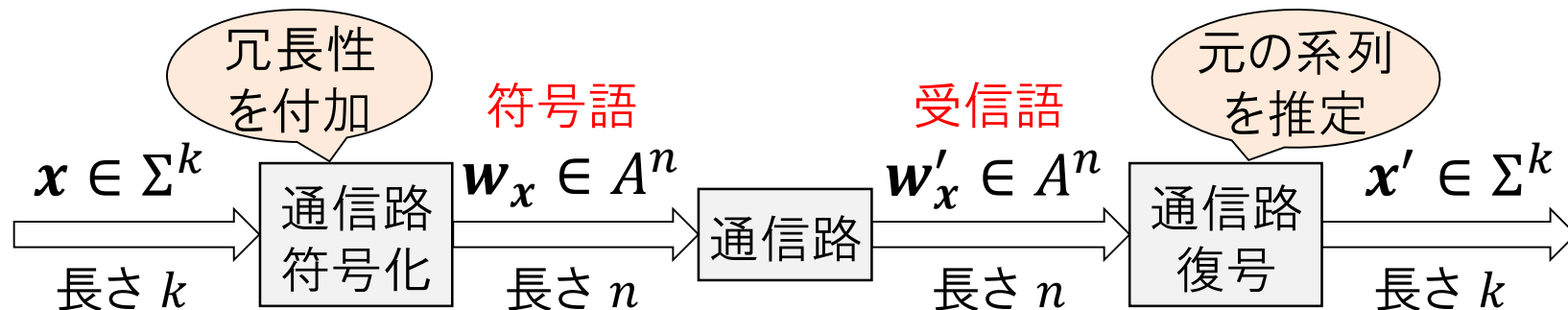


講義「情報理論」

第11回 通信路符号化の限界(2)

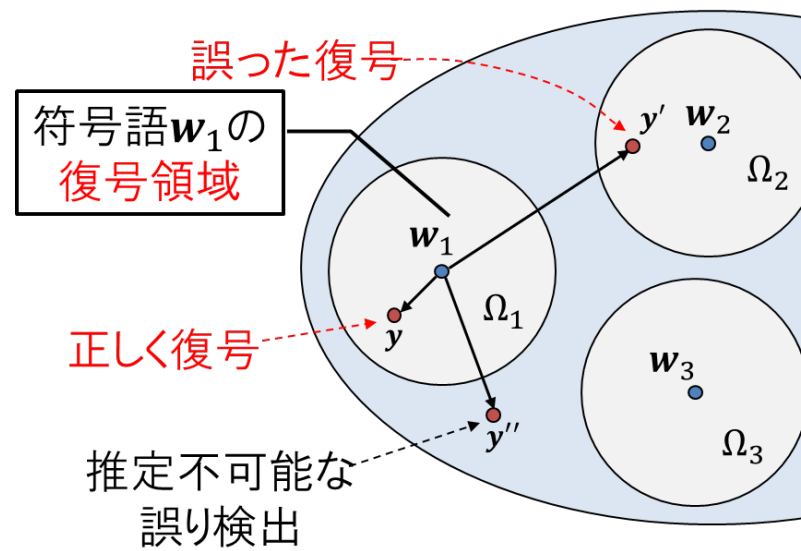
情報理工学部門 情報知識ネットワーク研究室
喜田拓也

通信路符号化の考え方（おさらい）



通信路符号化は、入力される q 元記号列を長さ k 毎に区切り、各ブロックに長さ n の等長な r 元記号列を割り当てることで行う

符号語どうしが受信空間(A^n)内で十分に離れていれば、受信語に少しの誤りが含まれていても、それに近い符号語へと推定できる



通信路容量（おさらい）

記憶のない定常通信路の通信路容量（定理7.1より）

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

（単位は、ビットあるいはビット/通信路記号）

2元対称通信路の通信路容量： $C = 1 - \mathcal{H}(p)$

p は
ビット誤り率

通信路に記憶がある場合の通信路容量

拡大情報源を考える。すなわち、長さ n の入力系列を X_n ，出力系列を Y_n とし、 P_{X_n} を X_n の確率分布とすれば、

$$C = \lim_{n \rightarrow \infty} \left[\max_{P_{X_n} \in \mathcal{P}^n} \left\{ \frac{1}{n} I(X_n; Y_n) \right\} \right].$$

誤り源のエントロピー
 $H(E)$ （ビット/記号）

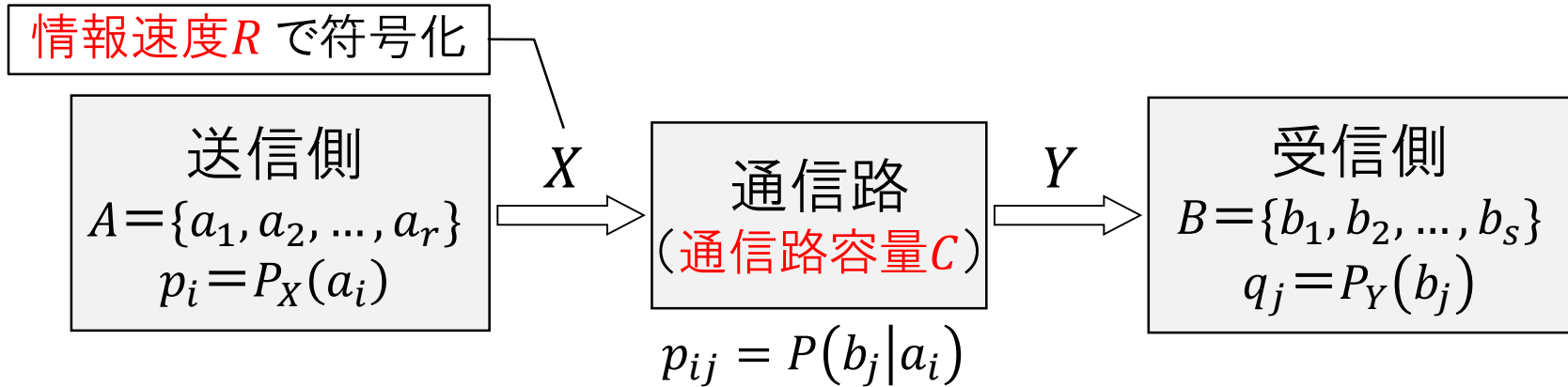
加法的2元通信路の通信路容量： $C = 1 - H(E)$

今日の内容

7.3 通信路符号化定理

7.4 通信システム全体としての情報伝達の限界

通信路容量と情報速度の関係



通信路を使って伝送できる、1記号あたりの情報量の最大値は、その通信路の**通信路容量 C** (ビット/記号)

ある通信路符号化により、通信路に入力される符号語の1記号あたりの平均情報量は、その符号の**情報速度 R** (ビット/記号)

情報速度 $R < \text{通信路容量 } C$ ならよさそうだ！

R を C と比べて、どのくらい低くすれば良いだろうか？
それで、どのくらい信頼性を向上できるのか？

通信路符号化定理(定理7.4)

定理7.4[通信路符号化定理]

通信路容量 C である通信路に対し、 $R < C$ であれば、情報速度 R の符号で復号誤り率がいくらでも小さいものが存在する。しかし、 $R > C$ であれば、そのような符号は存在しない。

※ Shannonの第2符号化定理とも呼ばれる

通信路容量を超えない情報速度でなら、
いくらでも精度よく通信できる
ような符号法がある！！



※でも具体的な符号の構成方法は分かっていない……

通信路符号化定理の意味するところ

ビット誤り率が0.1である2元対称通信路があるとする。

この通信路の通信路容量 C は

$$C = 1 - \mathcal{H}(0.1) \approx 0.531 \text{ (ビット/記号)}.$$

2元記号列を長さ k のブロックに区切って、長さ n の2元符号化する。
この符号の情報速度 R は $R = k/n$ となるが、定理より、

$$k/n < 0.531$$

となるような (n, k) に対し、ほとんど誤りなく情報を伝達できる符号が存在する。

つまり、元の系列の $n/k = 1/0.531 \approx 1.88$ 倍の長さになるように冗長性を付加すれば、10ビットに平均して1回起こるビット誤りを、ほとんどすべて訂正することができるような符号が存在する。

※ 実際の通信路のビット誤り率はもっと小さいので、必要な冗長性はごく小さい

通信路符号化定理の証明（前半の概要）

定理は、**符号の存在だけ保証**している。

Shannonはこれを**ランダム符号化**(random coding)を用いて証明。
ランダム符号化とは、受信空間から独立な無作為復元抽出を M 回繰り返すことにより M 個の符号語を選ぶ符号化法である。

証明は、ランダム符号化によって作られる符号 C の復号誤り率 $P_e(C)$ の**期待値** $E_C(P_e(C))$ を求め、情報速度 R と通信路容量 C に対して、 $R < C$ のとき符号長 n を $n \rightarrow \infty$ とすれば、 **$E_C(P_e(C)) \rightarrow 0$** となることを示すことによる。

これが示されれば、**復号誤り率が期待値以下となるような符号は必ず一つは存在する**。この事実から定理の前半は証明される。

※ 詳しくは教科書7.6節を参照

通信路符号化定理の証明（後半）

情報速度が $R > C$ の符号で、復号誤り率 P_e をいくらでも小さくできるものが存在すると**仮定する**.

すると、実際に、 R にいくらでも近い速度で情報を伝達することができる。

すなわち、通信路容量 C を超えた情報速度で情報を送ることができることになる。

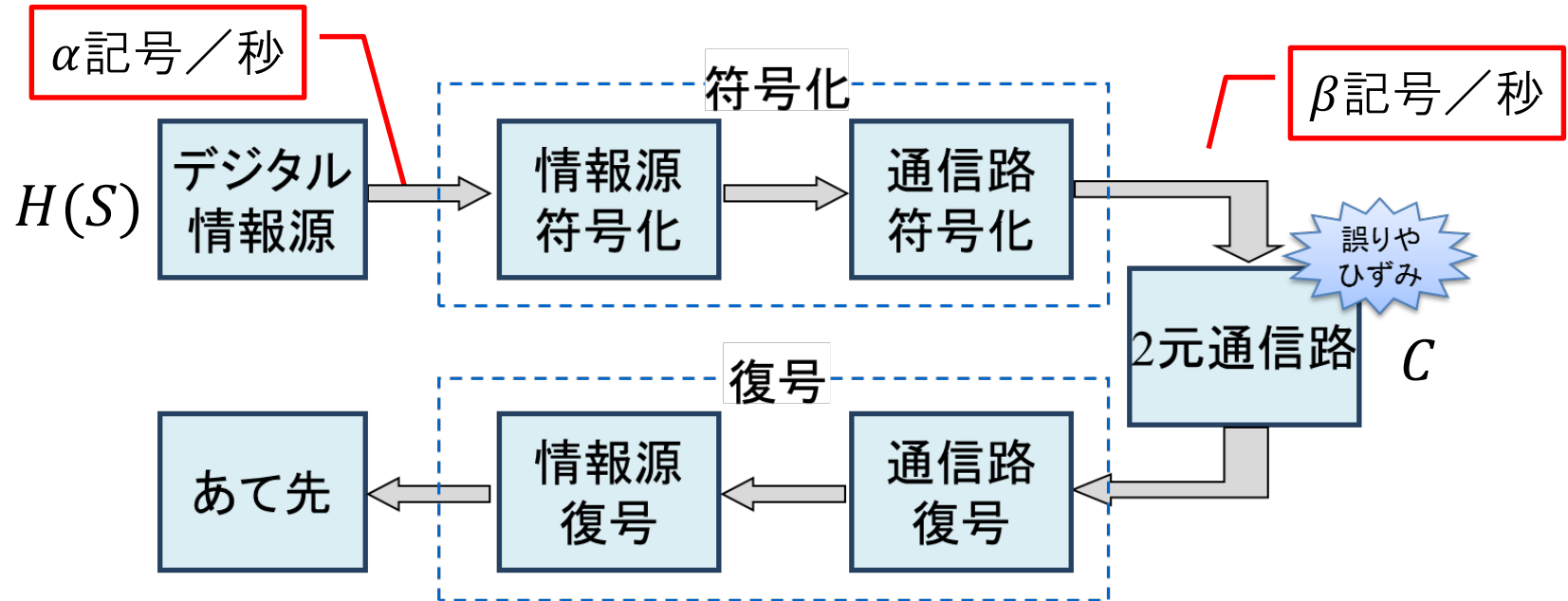
通信路容量の定義からこれは**不可能**である。

よって、そのような符号は存在しない。【証明終】

ちよつと休憩

もう一度，通信システムのモデル

与えられた情報源と通信路に対し**どれだけ良い通信ができるか？**



エントロピー $H(S)$ (ビット／情報源記号)の情報源 S ，
通信路容量 C (ビット／通信路記号)の通信路を考える

情報源から情報源記号が**毎秒 α 個発生**し，通信路では**毎秒 β 個**の
通信路記号が伝送されているとする。

この通信システムで伝送できる限界は？

このとき、情報源から発生する秒あたりの情報量は、

$$\mathcal{R} = \alpha H(S) \quad (\text{ビット/秒}).$$

一方、秒あたり通信路容量は

$$\mathcal{C} = \beta C \quad (\text{ビット/秒}).$$

したがって、

$$\mathcal{R} < \mathcal{C}$$

ならば任意に小さい誤り率で通信することができる。しかし、

$$\mathcal{R} > \mathcal{C}$$

の場合は、ひずみが生じてしまう。

$\mathcal{R} > \mathbf{C}$ の場合どのくらいひずむのか？

情報源 S の速度・ひずみ関数を $R(D)$ (ビット／情報源記号)とすると、平均ひずみが D 以下という条件の下で、秒あたりの情報量を

$$\mathcal{R}(D) = \alpha R(D)$$

まで落とすことができる。

このとき、 $\mathcal{R}(D_*) = \mathbf{C}$ を満たす D_* を考える。すると、 D_* よりも ε だけ多くの平均ひずみを許せば、情報速度

$$\mathcal{R}(D_* + \varepsilon) < \mathcal{R}(D_*) = \mathbf{C}$$

の符号化を作ることができる。

すなわち、情報源 S からの情報を、 D_* に任意に近い平均ひずみで送ることができる。

通信システム全体としての情報伝達の限界

定理7.6

情報速度 \mathcal{R} (ビット/秒)で発生する情報を通信路容量 \mathcal{C} (ビット/秒)の通信路を介して送るとき,

$$\mathcal{R} < \mathcal{C}$$

であれば, 任意に小さい誤り率で情報を伝送できる. また,

$$\mathcal{R} > \mathcal{C}$$

であれば, 情報源の速度・ひずみ関数が

$$\mathcal{R}(D_*) = \mathcal{C} \text{ (ビット/秒)}$$

を満たす D_* に対し, D_* に任意に近いひずみで情報を伝送できるが, D_* より小さい平均ひずみでは伝送できない.

例題7.3

1, 0 を 0.2, 0.8 の確率で発生する記憶のない情報源の出力系列を、ビット誤り率が0.1の2元対称通信路を通して送る。情報源は1秒に1記号を発生し、通信路も1秒に1記号を伝送する。このとき、

$$\mathcal{R} = \mathcal{H}(0.2) \approx 0.7219 \quad (\text{ビット/秒}),$$

$$\mathcal{C} = 1 - \mathcal{H}(0.1) \approx 0.5310 \quad (\text{ビット/秒}).$$

$\mathcal{R} > \mathcal{C}$ なので、**ひずみなしには通報を送れない。**

ひずみ測度として $d(x, y) = \begin{cases} 0; & x = y \\ 1; & x \neq y \end{cases}$ を用いる。

このとき、速度・ひずみ関数は右図の $p = 0.2$ の場合となる。これより $D_* = 0.0293$ が求まる。

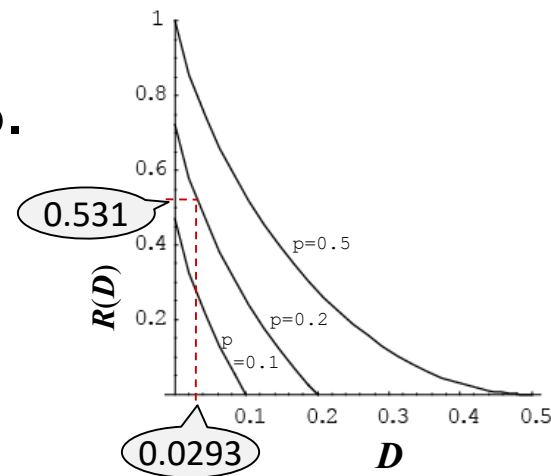
$$\because \alpha R(D_*) = \mathcal{C} \Rightarrow R(D_*) = 0.5310$$

$$\Rightarrow \mathcal{H}(0.2) - \mathcal{H}(D_*) = 0.5310$$

$$\Rightarrow \mathcal{H}(D_*) = 0.7219 - 0.5310 = 0.1909$$

$$\Rightarrow D_* = \mathcal{H}^{-1}(0.1909) = 0.0293.$$

すなわち、
 $\alpha = \beta = 1$



速度・ひずみ関数 $R(D)$

Try 練習問題7.3

今日のまとめ

通信路符号化定理

通信路符号化定理(Shannonの第2符号化定理)

ランダム符号化法による証明

符号語長が有限の場合の限界

信頼性関数

通信の限界

通信の理論的限界

次回

(具体的な)通信路符号化法について

符号長 n が有限のときの P_e の収束の速さ

符号長 n を無限大にすると、いくらでも復号誤り率を小さくできる。
 n が有限のとき、復号誤り率 P_e はどのように0に近づくだろうか？

定理7.5

記憶のない通信路に対し、ランダム符号化による符号長 n 、情報速度 R の符号 C の復号誤り率 P_e の期待値 $E_C(P_e(C))$ は、

$$E_C(P_e(C)) \leq 2^{-nE_r(R)}$$

を満たす。ただし、 $E_r(R)$ は次式で定義される。

$$E_r(R) = \max_{P_X, 0 \leq \rho \leq 1} [E_0(\rho, p) - \rho R]$$

ここに、 P_X は入力アルファベット上の分布であり、

$$E_0(\rho, p) = -\log_2 \sum_{y \in B} \left[\sum_{x \in A} P_X(x) P_{(Y|X)}(y|x)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

である。

ランダム符号化
指数

ρ は冗長度

Gallager関数

符号長 n , 情報速度 R の最適な符号法は？

系7.1

記憶のない通信路において, 復号誤り率 P_e が,

$$P_e \leq 2^{-nE_r(R)}$$

を満たす長さ n , 情報速度 R の符号が存在する.

では, 最も良い符号の収束速度はいくらなのか？

$P_e^*(n, R)$ を長さ n , 情報速度 R の符号の中での復号誤り率の最小値とする. それを達成する符号の収束速度 $E(R)$ は,

$$E(R) = \lim_{n \rightarrow \infty} -\log_2 P_e^*(n, R) / n$$

で与えられる. これは**信頼度関数** (reliability function) と呼ばれる. しかし, そのような符号の実際的な構成方法は知られていない.

LDPC(低密度パリティ符号): Robert G. Gallager (1963年)

ターボ符号: C. Berrou, A. Glavieux, P. Thitimajshima (1993年)

信頼度関数を計算するのは, 一般には難しい.

限界に迫る
符号化!

2元対称通信路に対する信頼度関数

p をビット誤り率とするとき、

$$p_1^* = \sqrt{p} / (\sqrt{p} + \sqrt{1-p})$$

となる p_1^* を求め、 $R_0 = 1 - \mathcal{H}(p_1^*)$ とおけば、 $0 \leq R \leq R_0$ となる情報速度 R に対しては

$$E(R) = 1 - R - 2 \log_2(\sqrt{p} + \sqrt{1-p}).$$

また、 $R_0 \leq R \leq C = 1 - \mathcal{H}(p)$ となる R に対しては、まず、

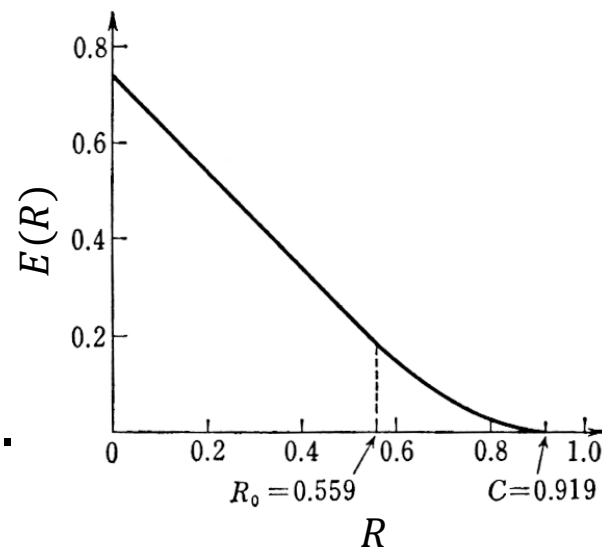
$$R = 1 - \mathcal{H}(p^*)$$

p^* は
 R の関数

を満たす p^* ($p < p^* < p_1^*$) を求め、

$$E(R) = p^* \log_2 \frac{p^*}{p} + (1 - p^*) \log_2 \frac{1-p^*}{1-p}$$

とすればよい。右図に、 $p = 0.01$ のときの $E(R)$ を示す。



2元対称通信路の
信頼度関数