

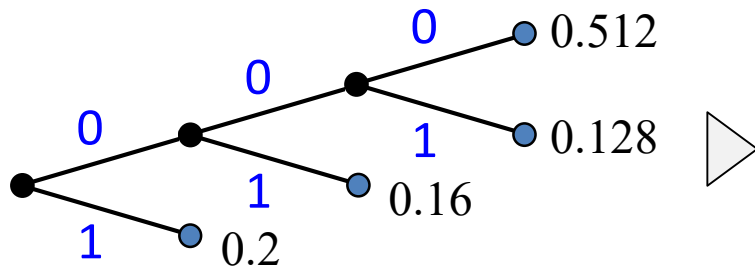
講義「情報理論」

第9回 通信路のモデル

情報理工学部門 情報知識ネットワーク研究室
喜田拓也

非等長情報源系列の符号化(おさらい)

1, 0を確率0.2, 0.8で発生する無記憶定常情報源 S を考える。
 S から発生する系列を4つ選び、ハフマン符号化を行う。



情報源系列を分割する分節木

各ブロックの平均長 \bar{n} は

$$\bar{n} = 1 \times 0.2 + 2 \times 0.16 + 3 \times 0.128 + 3 \times 0.512 = 2.44$$

右の符号の平均符号長 $L' = 1.776$

よって1記号あたりの平均符号長 L は

$$L = \frac{1.776}{2.44} = 0.728$$

情報源系列	確率	ハフマン符号
000	0.512	0
001	0.128	100
01	0.16	101
1	0.2	11

ひずみが許される場合の情報源符号化(おさらい)

定理 [ひずみが許される場合の情報源符号化定理]

平均ひずみ \bar{d} を D 以下に抑えるという条件の下で, 任意の正数 ε に対して, 情報源 S を1情報源記号あたりの平均符号長 L が

$$R(D) \leq L < R(D) + \varepsilon$$

となるような2元符号へ符号化できる. しかし, どのような符号化を行っても, $\bar{d} \leq D$ である限り, L をこの式の左辺より小さくすることはできない.

この定理は, 1情報源記号あたりの平均符号長を, 速度・ひずみ関数 $R(D)$ にいくらでも近づく符号化法の存在を示している

具体的な符号化方法はあるのか?

ひずみのない場合に比べてはるかに難しい!

教科書【例5.8】参照

今日の内容

6.1 通信路の統計的表現

6.2 記憶のない定常通信路

6.3 加法的2元通信路

もう一度，情報理論の問題について

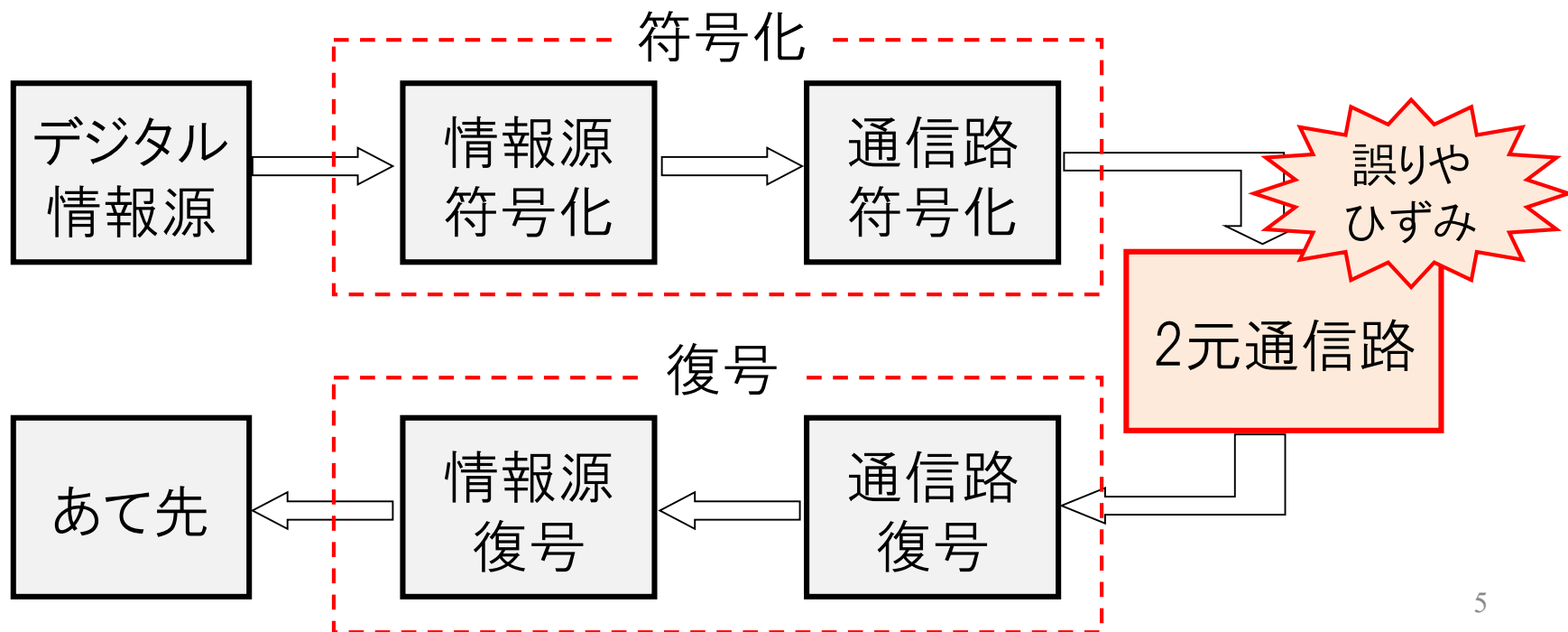
情報理論が取り組む4つの問題

【問題1】できるだけよい情報源符号化法(復号法)を見出すこと

【問題2】情報源符号化の限界を知ること

【問題3】できるだけよい通信路符号化法(復号法)を見出すこと

【問題4】通信路符号化の限界を知ること



通信路の統計的表現

雑音のある離散的通信路の定義:

時点毎に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,



$|A| = |B| = r$ のときは, **r 元通信路 (r -ary channel)** という

入力 X_0, X_1, \dots, X_{n-1} に対する Y_0, Y_1, \dots, Y_{n-1} の確率分布

$$P_{Y_0, Y_1, \dots, Y_{n-1} | X_0, X_1, \dots, X_{n-1}}(y_0, y_1, \dots, y_{n-1} | x_0, x_1, \dots, x_{n-1})$$

= $[X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}$ の条件の下で
 $Y_0 = y_0, Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}$ となる確率]

記憶のない定常通信路(memoryless channel)

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を、**記憶のない通信路**という

さらに、時間をずらしても統計的性質が変わらないとき、これを**記憶のない定常通信路**と呼ぶ

記憶のない定常通信路では、入力 X が通信路に投入されたときに出力 Y が出る条件付確率 $P_{Y|X}(y|x)$ が、すべての時点において同一である。したがって、

$$P_{Y_0 \dots Y_{n-1} | X_0 \dots X_{n-1}}(y_0, \dots, y_{n-1} | x_0, \dots, x_{n-1})$$

$$= \boxed{\phantom{P_{Y_0 \dots Y_{n-1} | X_0 \dots X_{n-1}}(y_0, \dots, y_{n-1} | x_0, \dots, x_{n-1})}}$$

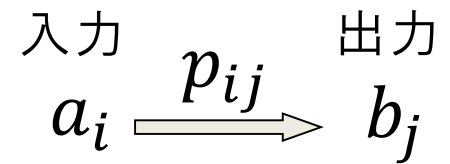
通信路行列と通信路線図

r 元入力アルファベット $A = \{a_1, a_2, \dots, a_r\}$,

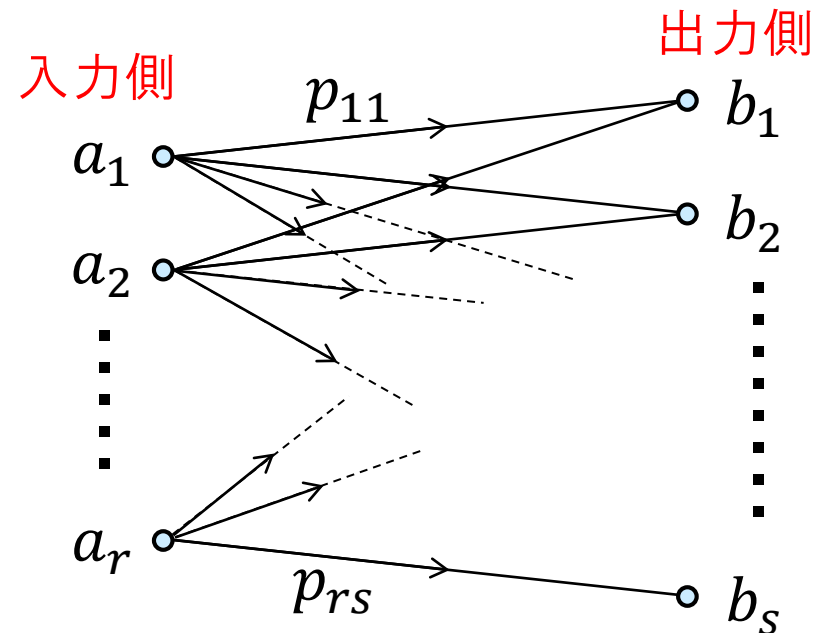
s 元出力アルファベット $B = \{b_1, b_2, \dots, b_s\}$,

入出力の関係が条件付確率 $p_{ij} = P_{Y|X}(b_j|a_i)$

で与えられる記憶のない定常通信路を考える



$$T = \begin{matrix} & \text{出力側} & & & \\ & & & & \\ \begin{matrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{matrix} & & & & \\ & & & & \text{入力側} \end{matrix}$$



p_{ij} を (i, j) 要素とする通信路行列

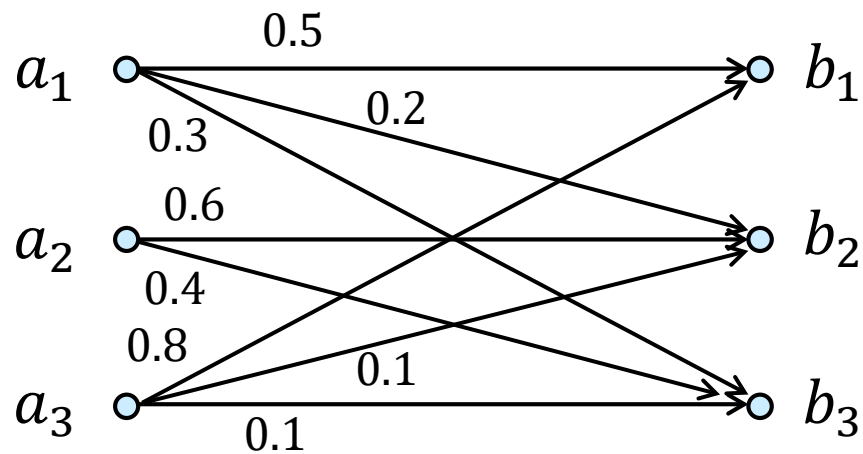
通信路線図

例題6.1

$P(x y)$		y		
		b_1	b_2	b_3
x	a_1	0.5	0.2	0.3
	a_2	0	0.6	0.4
	a_3	0.8	0.1	0.1

$$T = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0 & 0.6 & 0.4 \\ 0.8 & 0.1 & 0.1 \end{bmatrix}$$

通信路行列 T



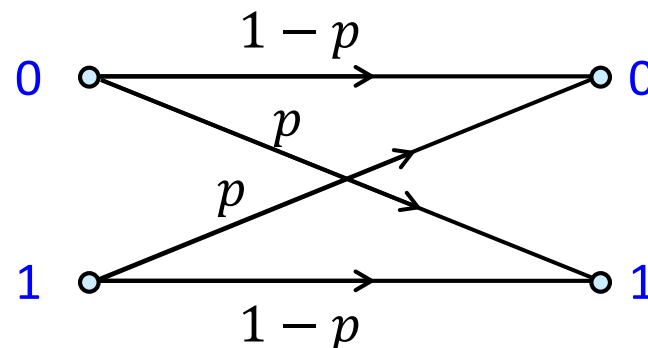
[Try 練習問題6.1](#)

一様な通信路

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

2重に一様

$$T = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \end{matrix} \begin{matrix} 0 \\ 1 \end{matrix}$$



2元対称消失通信路

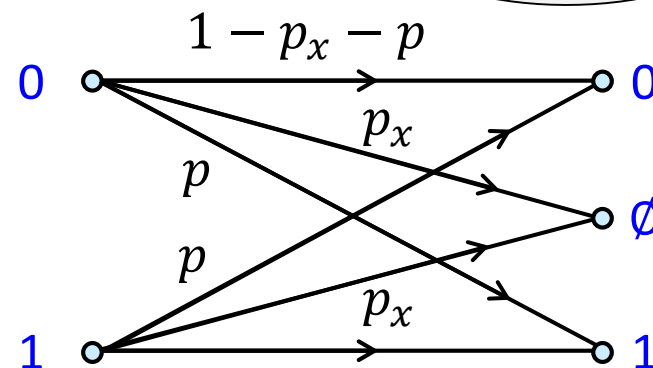
入力アルファベットは $\{0, 1\}$

出力アルファベットは $\{0, 1, \emptyset\}$

(\emptyset は消失を表現)

入力に対して一様

$$T = \begin{matrix} & \begin{matrix} 0 & \emptyset & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 1-p_x-p & p_x & p \\ p & p_x & 1-p_x-p \end{pmatrix} \end{matrix} \begin{matrix} 0 \\ 1 \end{matrix}$$



ちよつと休憩

加法的2元通信路

入力と出力のアルファベットが共に $\{0,1\}$ である2元通信路は、**誤りの有無**を用いて表すことができる

t 時点での誤りを確率変数 $E_t \in \{0,1\}$ で表すと、出力 Y_t は入力 X_t に誤り E_t を加えたものとみなせる

$$Y_t = X_t \oplus E_t,$$

$$E_t = \begin{cases} 0 & \text{誤りなし} \\ 1 & \text{誤り発生} \end{cases}$$

入力 0 → 出力 0

入力 0 → 出力 1 $0 \oplus 1 = 1$

入力 1 → 出力 0 $1 \oplus 1 = 0$

入力 1 → 出力 1

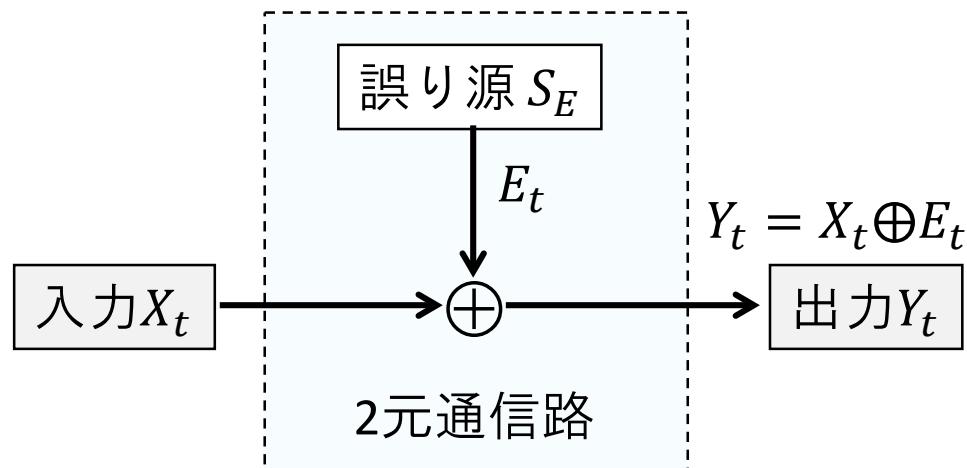


図6.3 加法的2元通信路モデル

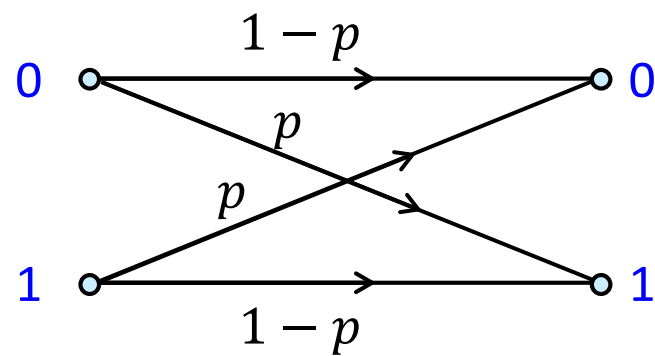
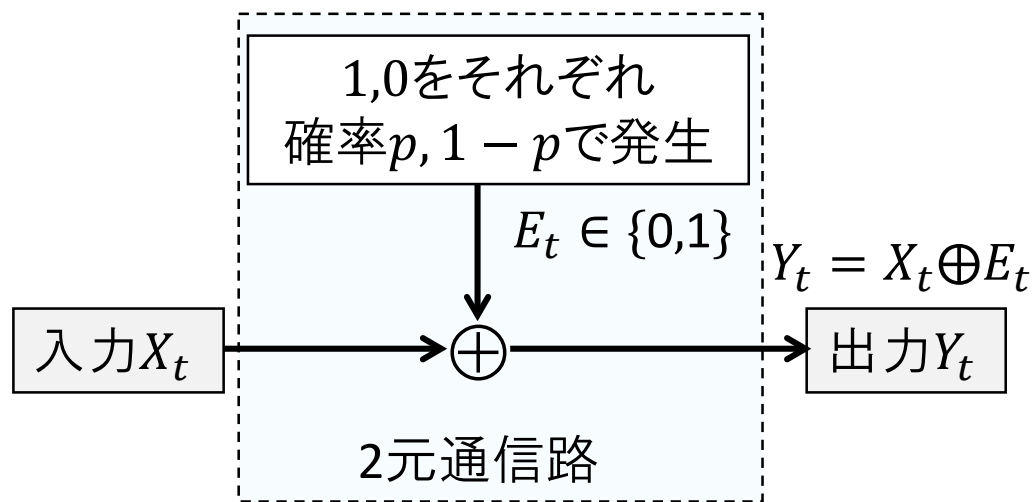
※誤りの発生は入力と統計的に独立であると仮定される

ランダム誤り通信路

加法的2元通信路の誤り源 S_E が、1,0をそれぞれ確率 $p, 1 - p$ で発生させる記憶のない定常2元情報源とする. このとき、0から1への誤りも、1から0への誤りも、他の時点の入出力とは無関係に確率 p で発生する. これは**2元対称通信路**に他ならない

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



2元対称通信路の通信路線図

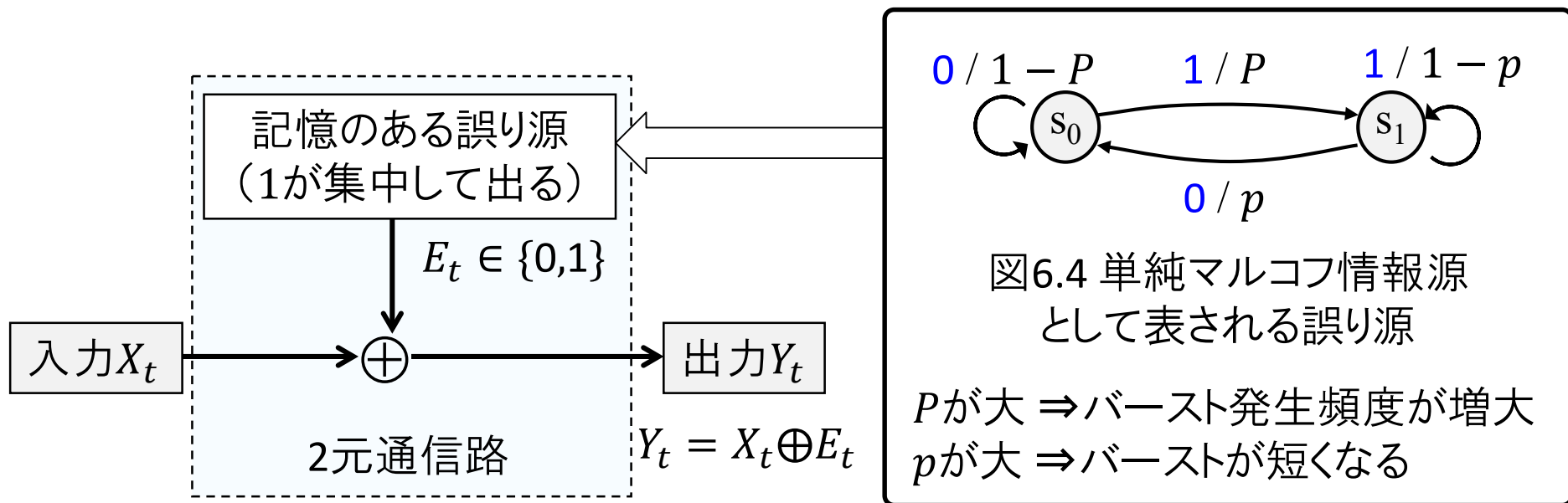
バースト誤り通信路

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モデル(誤り源に記憶がある代表的なモデル)

密集して生じる誤りをバースト誤り(burst error)と呼ぶ

例えば、誤り源から発信される系列が次のようになる

0000000**1111111**0000**1111**0000・・・ (ソリッドバーストの例)



ソリッドバースト誤りの平均長

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 $P_L(\ell)$ を求めると,

$$P_L(\ell) = \underline{(1-p)^{\ell-1}} \underline{p} \quad (\ell = 1, 2, \dots)$$

となる

最初の1の後に,
1が $\ell - 1$ 連続する確率

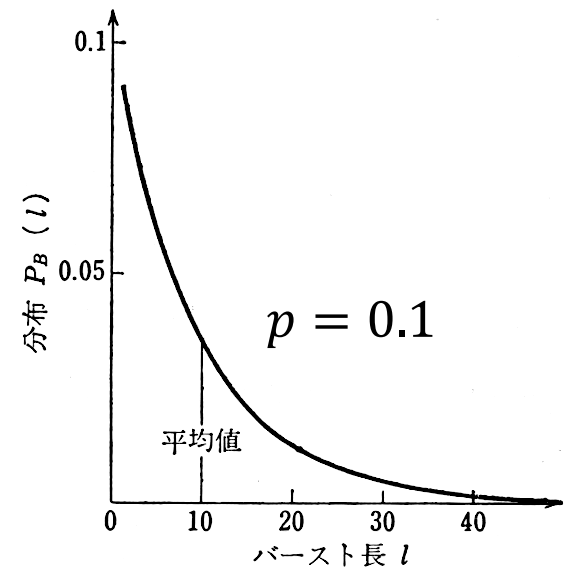
最後に0が出る確率

.....00 ^{ℓ} 1111100.....

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

$$\begin{aligned}\bar{\ell} &= \sum_{\ell=1}^{\infty} \ell P_L(\ell) \\ &= p \sum_{\ell=1}^{\infty} \ell (1-p)^{\ell-1} \\ &= \frac{1}{p}\end{aligned}$$

手計算で求まります
ノート6.1



バースト長の分布の例

ソリッドバースト誤り源のビット誤り率(例題6.2)

図6.4の誤り源の状態遷移行列 Π は

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

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

$$w_0 = \frac{p}{P + p}, \quad w_1 = \frac{P}{P + p}.$$

よって, 誤り源の出力 E が 1 となる確率 $P_E(1)$ を求めると

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

$$= \frac{P}{P + p}.$$

ビット誤り率

Try 練習問題6.2

その他のバースト誤りモデル

ギルバートモデル (Gilbert model)

正誤が混在するバースト誤り
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
バースト長の期待値は $1/p$

$$\text{ビット誤り率} \frac{Ph}{P+p}$$

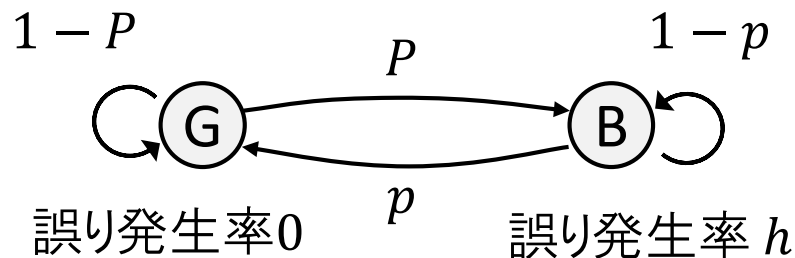
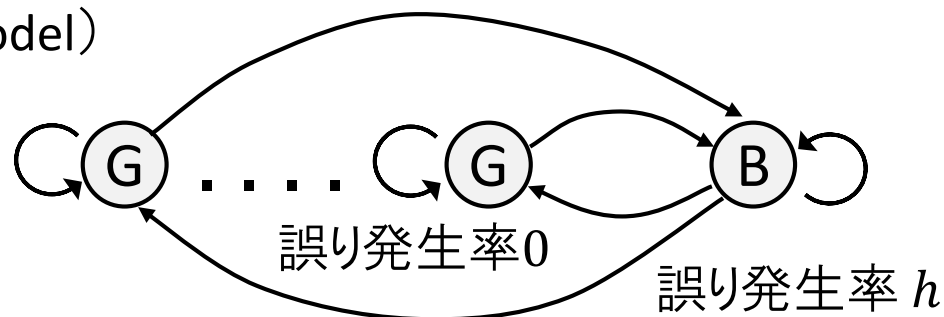


図6.6 ギルバートモデル

フリッチマンモデル (Fritchman model)

ギルバートモデルの
良状態を増やしたもの



フリッチマンモデル

今日のまとめ

6.1 通信路の統計的表現

6.2 記憶のない定常通信路

通信路行列と通信路線図

2元対称通信路(2重に一様な通信路)

2元対称消失通信路(入力に対して一様な通信路)

6.3 加法的2元通信路

ランダム誤り通信路 = 2元対称通信路

バースト誤り通信路(ソリッドバースト)

ギルバートモデル, フリッチマンモデル

次回

通信路符号化の限界に関する理論

通信路にまつわる各種エントロピー

入力記号側のエントロピー:

$$H(X) = - \sum_{i=1}^r p(a_i) \log_2 p(a_i)$$

出力記号側のエントロピー:

$$H(Y) = - \sum_{j=1}^s q(b_j) \log_2 q(b_j)$$

条件付エントロピー:

$$H(X|Y) = - \sum_{i=1}^r \sum_{j=1}^s q(b_j) p(a_i|b_j) \log_2 p(a_i|b_j)$$

結合エントロピー:

$$H(X, Y) = - \sum_{i=1}^r \sum_{j=1}^s p(a_i, b_j) \log_2 p(a_i, b_j)$$

※入力 X の記号 a_i ($i = 1, \dots, r$)の生起確率 $p(a_i)$, 出力 Y の記号 b_j ($j = 1, \dots, s$)の生起確率 $q(b_j)$

(付録) Tunstall-Huffman 符号の効率

練習問題5.3の条件のとき，タンストール木の大きさ(使う符号語の数 N)を $N = 5$ としたときより $N = 6$ とするほうが，1記号あたりの平均符号語長が長くなる？

$N = 5$ のとき：

平均ブロック長 $\bar{n}_5 = 2.36$ ，ブロックあたりの平均符号長 $\bar{\ell}'_5 = 2.304$
よって，1記号あたりの平均符号長 $\bar{\ell}_5 \cong 0.97627$

$N = 6$ のとき：

平均ブロック長 $\bar{n}_6 = 2.60$ ，ブロックあたりの平均符号長 $\bar{\ell}'_6 = 2.544$
よって，1記号あたりの平均符号長 $\bar{\ell}_6 \cong 0.97846$



< あれえ？

(付録) Tunstall-Huffman 符号の効率

実は、そういうことはありうる！！

何故なら、ハフマン符号が(ブロック毎に)2元符号化しているから
(例えば3.4個分の長さの符号語とか作れないから)

Tunstall木を大きくすると平均ブロック長は長くなるが、ハフマン符号化したときに1記号あたりの平均符号長が短くなるとは限らない
2001年に理論的な解析結果が発表されていました！

Serap A. Savari and Wojciech Szpankowski, "On the Analysis of Variable-to-Variable Length Codes" (2002年に同タイトルのショートペーパーが*IEEE International Symposium on Information Theory*で発表されている)

上記論文内の定理2:

$$\limsup_{M \rightarrow \infty} \log_2 M \cdot R_{T-H}(M) \leq \mathcal{H} \log_2 \left(\frac{2 \log_2 e}{e} \right)$$

M は符号語の数
 $R_{T-H}(M)$ は符号の冗長度

\mathcal{H} は情報源のエントロピー
 $\log_2((2 \log_2 e)/e) \cong 0.086$

