7. 相互情報量と通信路モデル

通信路モデルは相互情報量概念に基づいて構成される．簡単な例題を用いて情報量の定量的尺度について復習したあと，相互情報量概念を導入し，その性質を調べる．これらの基礎に基づいて通信路モデルを導入する．

7.1 情報の定量的尺度

情報の量的側面がどのように定式化されるか，6 章で使った投球の例を用いて説明しよう．6 章と同様に，投手の球種は，直球と変化球だけとし，直球の確率をpとすると，変化球の確率は$1-p$となる．はじめに，この投手の球種は状況に関わらず決まるものと仮定すれば，この投手の投球内容は，確率pで情報源記号「直球」を発生し，確率$1-p$で情報源記号「変化球」を発生する，記憶のない情報源Sとしてモデル化できる（図 1）．

これまでの章でみてきたように，2元記号を使ってこの情報源の符号化を試みると，その平均符号長の下限をSの1次エントロピー$H_1(S) = -(p\log_2 p + (1-p)\log_2(1-p))$に限りなく近づけることができるが，$H_1(S)$未満にはできない．この情報源平均符号長の下限は次のように「サプライズ」（まれなことが起きたことに対する驚き）として捉えることもできる．

確率pで発生する事象の起きたことを伝えるメッセージのサプライズは，pが小さいほど大きい．このことを反映してそのメッセージのサプライズを，$-\log_2 p$ビットであるとし，以下では，そのメッセージの情報量と呼ぶ．$p=\frac{1}{2}$のときだと，そのメッセージの情報量は1ビットである．メッセージの情報量は，メッセージの生起確率pがゼロに近くづくにつれて増え，逆に1に近づくにつれて減る．

確率pで情報源記号「直球」が発生し，確率$1-p$で情報源記号「変化球」を発生する情報源から発生する情報源記号が何であったかを伝えるメッセージの平均情報量は，

$$H_1(S) = -(p\log_2 p + (1-p)\log_2(1-p))\text{ビット}$$

である．

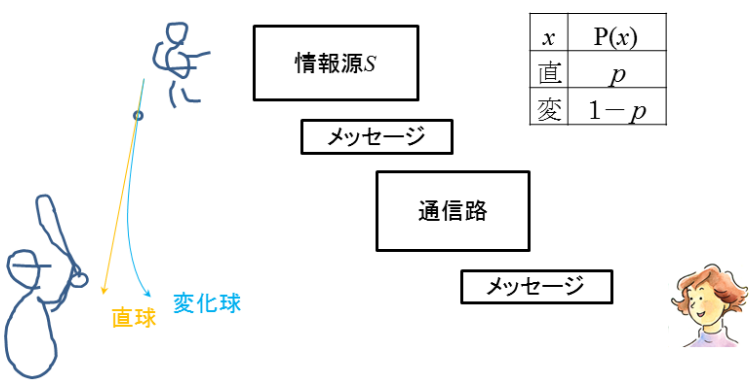

図 1：投球の例題

$H_1(S)$の値は，$p=\frac{1}{2}$のとき最も大きい．つまり，直球の生起確率と変化球の生起確率が等しい時に最大値1ビットになる．$p=\frac{1}{4}$の場合は，メッセージは$\frac{1}{4}$の確率で$\log_2 4 = 2$ビットの情報をもたらすが，$\frac{3}{4}$の確率で，$-\log_2\frac{3}{4} \approx 0.415$ビットだけの情報しかもたらさない．両者を平均すれば，メッセージが平均的にもたらす情報量は，

$$-\left(\frac{1}{4}\log_2\frac{1}{\frac{1}{4}}+\frac{3}{4}\log_2\frac{1}{\frac{3}{4}}\right) \approx \left(\frac{1}{4}\times 2+\frac{3}{4}\times 0.415\right) \approx 0.811 \text{ ビット}$$

となる．$p=\frac{1}{8}$の場合は，メッセージが平均的にもたらす情報量は約0.543ビット，$p=0.001$の場合は，約0.0114ビットとなる．

以下では，

$$\mathcal{H}(x) \equiv x\log_2\frac{1}{x}+(1-x)\log_2\frac{1}{1-x}$$

と表記する．$\lim_{x\downarrow 0}(x\log_2 x)=0$に注意すれば，エントロピー関数$\mathcal{H}(x)$の形状は，図 2 のようになる．

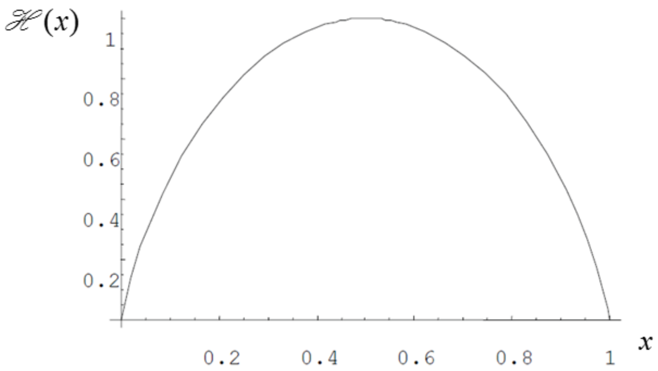

図 2: エントロピー関数$\mathcal{H}(x) \equiv x\log_2\frac{1}{x}+(1-x)\log_2\frac{1}{1-x}$の形状

◎ なぜ log か？

生起確率pの事象が起きたことを知らせる「文」の情報量を$I(p)$と一般的に表記しよう．これまでの議論では，$I(p)=C\log_e p$としてきた．なぜそのようにしたのか？

ここでは，$I(p)$を定めるために log という関数を用いたことが恣意的なものではなく，自然なものであることを示そう．そのために，$I(p)$の定義を決めるときに用いられた（はずの）原則について再考してみよう．

1. 「文」の情報量は「文」の価値を表すものと考える．「文」の価値が高いほど，その「文」の情報量も大きいとする．
2. 珍しい事象の生起を告げる「文」ほどその価値が高い．

3. 事象が珍しいほど，それを告げるためには長い「文」が必要になる．
4. 事象を伝えるために必要な「文」の長さの最小値をその「文」の価値とする．
5. 事象Aと事象Bが独立であるとき，事象Aと事象Bの両方が起きたことを告げる「文」は，概ね「Aが起き，Bも起きた」となるので，その長さの最小値は，事象Aが起きたことを告げる「文」（「Aが起きた」）の長さの最小値と，事象Bが起きたことを告げる「文」（「Bが起きた」）の長さの最小値の和である．
6. $I(p)$の関数はpの微分可能関数にしておきたい．

第 5，6 項から$I(p)$が$C\log_e p$という関数であることが決まる．つまり，事象Aの生起確率をp，事象Bの生起確率をqとすると，事象Aと事象Bが独立であるとき，事象Aと事象Bの両方が起きる確率は$p \times q$であるので，第 5 項は，

$$I(p \times q) = I(p) + I(q)$$

であることを要請している．このような条件を満足する微分可能な関数は，$I(p) = C\log_e p$しかない（Cは定数）．

【演習】上記を確かめてみよう．

生起確率pのメッセージのもつ情報量を単に$\frac{1}{p}$にせず，それにlogをつけて$\log_2 \frac{1}{p}$と定義する理由は，直観的には次のようにも考えられる．一般に，xが与えられたとき，$\log_2 x$はxを2進数表示したときの桁数を表す．生起確率pのメッセージの生起確率が2^{-10}であったとき，そのメッセージを2進数を用いて表示するのであれば，どのような2進数表示を割り当てたらいいだろうか？

そのようなメッセージが他にもあり，それぞれに異なる2進数表示を割り当てなければならないとすると，生起確率が2^{-10}のメッセージには10桁の2進数表示を割り当てるのが素直であるように思える．というのは，他のメッセージもすべて同じく生起確率が2^{-10}とすれば，そのようなメッセージは全部で2^{10}個あることになり，それらを区別するためには，10桁の2進数表示が必要であるからだ．他のメッセージのなかに生起確率が2^{-3}のメッセージがあるとすれば，それを表示するために，そのために10桁の2進数表示2^7個を確保して，その2^7個の10桁の2進数表示をひとまとめにして，$\log_2 \frac{2^{10}}{2^7} = 10 - 7 = 3$桁の2進数表示にしてしまう．このように，生起確率$p$のメッセージに長さ$\log_2 \frac{1}{p}$の2進数表示を割り当てるようにすると，すべてのメッセージに対する2進数表示の平均の桁数（＝長さ）は，そうでない割り当てをした時よりも必ず小さくなることが示される．つまり，各メッセージに対して異なる2進数表示を割り当てるという条件を課す限り，それよりも平均桁数を減らすことはできな

いという意味で，下限を与える．シャノンは確率モデルの下で発生するメッセージに対して有限個の記号を用いて異なる表示を割り当てたとき，その表示の平均長の下限を与える量として，情報量として規定した．このことはクラフトの不等式，情報源符号化定理で学んだとおりである．

今度は，知りたいことについての直接的な答えが得られない場合について考えてみよう．投球の例でいえば，投球内容そのものを伝えるメッセージは得られないが，前の投球内容や，投球前に行われた投手のしぐさ，さらには，投球後の観客のざわめきのように，投球内容に関わるメッセージが利用可能であり，その投球内容との関わりを表 1 のような結合確率モデルで与えられるとしよう．

表 1：情報源記号｛直球，変化球｝に対する結合確率$p(x,y)$の例

	前を見た	横を見た
直球	0.4	0.1
変化球	0.2	0.3

つまり，

- 投球時に前を見たとき，直球を投げる確率は 0.4
- 投球時に前を見たとき，変化球を投げる確率は 0.2
- 投球時に横を見たとき，直球を投げる確率は 0.1
- 投球前に横を見たとき，変化球を投げる確率は 0.3

によって示すことが可能であるとする．

上記の確率モデルからいろいろな確率が導ける．そもそも，直球が投げられる確率，変化球が投げられる確率は，それぞれ，

$$P(\text{直}) = P(\text{直},\text{前}) + P(\text{直},\text{横}) = 0.4 + 0.1 = 0.5$$
$$P(\text{変}) = P(\text{変},\text{前}) + P(\text{変},\text{横}) = 0.2 + 0.3 = 0.5$$

である．

一方，投球前に前を見ていたか，横を見ていたかについては，

$$P(\text{前}) = P(\text{直},\text{前}) + P(\text{変},\text{前}) = 0.4 + 0.2 = 0.6$$
$$P(\text{横}) = P(\text{変},\text{横}) + P(\text{変},\text{横}) = 0.1 + 0.3 = 0.4$$

となる．さらに，前を見ていたという答えを得たという条件下で，直球が投げられるか，変化球が投げられるかについては，

$$P(\text{直}|\text{前}) = \frac{P(\text{直},\text{前})}{P(\text{直},\text{前}) + P(\text{変},\text{前})} = \frac{0.4}{0.4 + 0.2} = \frac{2}{3}$$
$$P(\text{変}|\text{前}) = \frac{P(\text{変},\text{前})}{P(\text{直},\text{前}) + P(\text{変},\text{前})} = \frac{0.2}{0.4 + 0.2} = \frac{1}{3}$$

であることが，結合確率モデルから導ける．同様に，

$$P(\text{直}|\text{横}) = \frac{P(\text{直},\text{横})}{P(\text{直},\text{横}) + P(\text{変},\text{横})} = \frac{0.1}{0.1+0.3} = \frac{1}{4}$$
$$P(\text{変}|\text{横}) = \frac{P(\text{変},\text{横})}{P(\text{直},\text{横}) + P(\text{変},\text{横})} = \frac{0.3}{0.1+0.3} = \frac{3}{4}$$

である．一方，直球か変化球かがわかった場合については，

$$P(\text{前}|\text{直}) = \frac{P(\text{前},\text{直})}{P(\text{前},\text{直}) + P(\text{横},\text{直})} = \frac{0.4}{0.4+0.1} = \frac{4}{5}$$
$$P(\text{横}|\text{直}) = \frac{P(\text{横},\text{直})}{P(\text{前},\text{直}) + P(\text{横},\text{直})} = \frac{0.1}{0.4+0.1} = \frac{1}{5}$$
$$P(\text{前}|\text{変}) = \frac{P(\text{前},\text{変})}{P(\text{前},\text{変}) + P(\text{横},\text{変})} = \frac{0.2}{0.2+0.3} = \frac{2}{5}$$
$$P(\text{横}|\text{変}) = \frac{P(\text{変},\text{横})}{P(\text{前},\text{変}) + P(\text{横},\text{変})} = \frac{0.3}{0.2+0.3} = \frac{3}{5}$$

となる．

この状況で，「直球か変化球かを知りたい」という問いに対する，「前を向いていたか，横を向いていたか」という必ずしも直接的ではない，回答を得ることで，どれくらいのゲインがあると思えばいいだろうか？

この問いに答えるために，次の 2 つの量を計算してみよう．第一は，何も教えてもらっていない状況で，直球か変化球かを教えてもらったときに得る平均情報量．第二は，前を見ていたか，横を見ていたかを教えてもらったあとで，直球か変化球かを教えてもらったときに得る平均情報量である．

第一の場合は，直球であることを告げるメッセージがもたらす情報量は，

$$\log_2 \frac{1}{P(\text{直})} \quad \text{ビット}$$

である．同様に，変化球であることを告げるメッセージのもたらす情報量も1ビットであり，直球か変化球かを告げるメッセージからは，平均すると1ビットの情報量が期待される．

第二の，前を見ていたか，横を見ていたかを教えてもらった場合については次のようになる．「前を見ていた」という報告を受けたとすれば，直球の確率は$\frac{2}{3}$，変化球の確率は$\frac{1}{3}$であるので，直球か変化球のいずれであるかを告げるメッセージからは，平均すると

$$\frac{2}{3}\log_2 \frac{1}{\frac{2}{3}} + \frac{1}{3}\log_2 \frac{1}{\frac{1}{3}} \approx 0.918 \quad \text{ビット}$$

の情報量が期待される．一方，「横を見ていた」という報告を受けたとすれば，直球の確率は$\frac{1}{4}$，変化球の確率は$\frac{3}{4}$であるので，直球か変化球のいずれであるかを告げるメッセージからは，平均すると

$$\frac{1}{4}\log_2 \frac{1}{\frac{1}{4}} + \frac{3}{4}\log_2 \frac{1}{\frac{3}{4}} \approx 0.811 \quad \text{ビット}$$

の情報が得られる．そもそも「前を見ていた」という報告が生じる確率は$\frac{3}{5}$,「横を見ていた」という報告が生じる確率は$\frac{2}{5}$であるので，直球か変化球のいずれであるかという問いに対するメッセージから期待される情報量をその割合で平均すると，

$$\frac{3}{5} \times 0.918 + \frac{2}{5} \times 0.811 \approx 0.8755 \quad \text{ビット}$$

となる．

何も教えてもらわなかったときに直球か変化球のいずれであるかを告げるメッセージから期待される情報量は1ビットであったから，「前を見ていた」か「横を見ていた」かを教えてもらうことで，直球か変化球のいずれであるかを告げるメッセージによって平均的に得られる情報量は，

$$1 - 0.8755 = 0.1245 \quad \text{ビット}$$

だけ減ったことになる．これがまさに，「前を見ていたか，横を見ていたか」という問いに対する直接的でない回答が，平均的にもたらす情報量である．

換言すると，直球か変化球のいずれであるかを告げるメッセージを発生する情報源は，当初は1ビットの不確かさをもっていたが，前を見ていたか横を見ていたか教えられると，たとえそれが直接的でなくても，平均的に約0.8755ビットの不確かさに減少する．そして，減少した0.1245ビットは，まさしく前を見ていたか横を見ていたかを告げるメッセージの情報量である．

問い　直球か変化球のいずれであるかを伝えてくれるメッセージが，前を見ていたか，横を見ていたかについて，不確かさを平均的にどれだけ減らすか，算出せよ．

問い　$P(\text{前},\text{直}), P(\text{前},\text{変}), P(\text{横},\text{直}), P(\text{横},\text{変})$の値を変えたとき，直球と変化球のいずれであるかを告げるメッセージがもたらす情報量はどう変化するか？

7.2 相互情報量

2個の確率変数X,Yに対して，結合確率$P(X,Y)$が表 2 のように定義できるとしよう．このような状況では，二つの確率変数の間で一方の生起を知ると，他方の生起についてヒントが得られる．最悪の場合，まったくノーヒントのこともあるが，一方の生起を知って，情報を失うことはない．

X,Yの間の相互情報量は，一方の値を知ったとき減少するエントロピー：

$$I(X;Y) = H(Y) - H(Y \,/\, X) = H(X) - H(X|Y)$$

として定義される．ここで，$H(X|Y) = -\sum_x \sum_y P(x,y) \log_2 P(x|y)$である．2番目の等号が成立することはただちに自明ではないが，以下に示すように，相互情報量は，確率変数の間の対称的な概念である．

表 2.　2個の確率変数X,Yに対する結合確率$P(X,Y)$

$P(X,Y)$		Y: y_1	$\cdots$	y_s	$P(x)$
X	x_1	$P(x_1,y_1)$	$\cdots$	$P(x_1,y_s)$	$P(x_1)=\sum_j P(x_1,y_j)$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	x_r	$P(x_r,y_1)$	$\cdots$	$P(x_r,y_s)$	$P(x_r)=\sum_j P(x_r,y_j)$
$P(y)$		$P(y_1)=\sum_i P(x_i,y_1)$	$\cdots$	$P(y_s)=\sum_i P(x_i,y_s)$	1

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= -\sum_x P(x)\log_2 P(x) + \sum_x\sum_y P(x,y)\log_2 P(x|y) \\
&= -\sum_x\sum_y P(x,y)\log_2 P(x) + \sum_x\sum_y P(x,y)\log_2 P(x|y) \\
&= \sum_x\sum_y P(x,y)\log_2\frac{P(x|y)}{P(x)} \\
&= \sum_x\sum_y P(x,y)\log_2\frac{P(x,y)}{P(x)P(y)}
\end{aligned}
$$

このように，$H(X)-H(X|Y)$はXとYについて対称な式となるから，$H(Y)-H(Y|X)$と等しいことがわかる．さらに，$-\sum\sum P(x,y)\log_2 P(x,y)$と置くと，

$$I(X;Y)=I(Y;X)=H(X)+H(Y)-H(X,Y)$$

となる．

6つの量：$H(X), H(Y), H(X,Y), H(X|Y), H(Y|X), I(X;Y)=I(Y;X)$の間の関係は図 3 に示したベン図と同等の図式で捉えることができる．

7.3　相互情報量の性質

相互情報量には次の性質がある．

$$0 \le I(X;Y) \le \min\{H(X), H(Y)\}$$

(1)　$0 \le I(X;Y)$の証明：

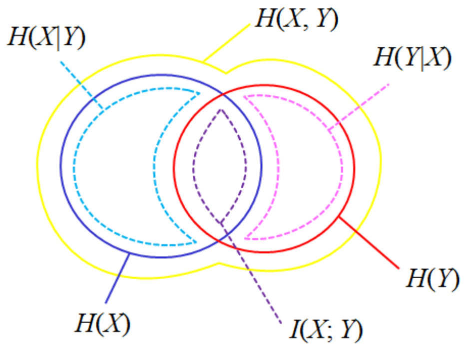

図 3.　$H(X), H(Y), H(X,Y), H(X|Y), H(Y|X), I(X;Y) = I(Y;X)$の間の関係

$$
\begin{aligned}
I(X;Y) &= \sum_x \sum_y P(x,y) \log_2 \frac{P(x,y)}{P(x)P(y)} \\
&= \sum_x \sum_y P(x,y) \log_2 P(x,y) - \sum_x \sum_y P(x,y) \log_2 P(x)P(y)
\end{aligned}
$$

ここで，

$$
\sum_{x,y} P(x)P(y) = \sum_x P(x) \left(\sum_y P(y) \right) = \sum_x P(x) = 1
$$

であるので，以前導入したシャノンの補助定理：

任意の$p_1, p_2, \cdots, p_M, p_i \geq 0, \sum_i p_i = 1$に対して，$q_1, q_2, \cdots, q_M$を

$q_1 + q_2 + \cdots + q_M \leq 1$，ただし，$p_i \neq 0$ ならば$q_i \neq 0$

なる非負の数とすれば，

$$
-\sum_i p_i \log_2 q_i \geq -\sum_i p_i \log_2 p_i = H_1(S)
$$

が成立する．ここで，等号は，$p_i = q_i \ (i = 1, \cdots, M)$のときそのときに限り成立する．

から

$$
-\sum_{x,y} P(x,y) \log_2 P(x)P(y) \geq -\sum_{x,y} P(x,y) \log_2 P(x,y)
$$

が導けるので，

$$
I(X;Y) = \sum_x \sum_y P(x,y) \log_2 P(x,y) - \sum_x \sum_y P(x,y) \log_2 P(x)P(y) \geq 0
$$

が示せる．ここで，等号が成立するのは，

$$P(x,y)=P(x)P(y)$$

つまり，XとYが独立であるときである．　■

(2) $I(X;Y)\le \min\{H(X),H(Y)\}$の証明：

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

ここで，$H(Y|X)$は非負だから，$I(X;Y)\le H(Y)$である．
同様に，$I(X;Y)\le H(X)$が導ける．　■

7.4 通信路モデル

通信路は，入力アルファベット列$x_0,\cdots,x_{n-1}$が与えられたとき出力アルファベット列$y_0,\cdots,y_{n-1}$が生起する確率を規定する条件付き確率

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

の集まりで規定される．

以下では，通信路は，記憶のない定常通信路であると仮定する．ここで，

記憶のない通信路：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，それ以外の時点の出力にも入力にも独立である

記憶のない定常通信路：記憶のない通信路が定常である，つまり，時間をずらしても統計的性質が変わらない

この場合，次の式が成立する．

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

記憶のない定常通信路の表現：

$$T=\begin{bmatrix}p_{11} & \cdots & p_{1s}\\ \vdots & \ddots & \vdots\\ p_{r1} & \cdots & p_{rs}\end{bmatrix}\quad ただし，p_{ij}=P(b_j|a_i)$$

を通信路行列という．

通信路行列をビジュアルに表現するために，図 4 のような通信路線図が用いられる．
入力アルファベットの生起確率を

$$p=(p_1,p_2,\cdots,p_r)$$

とすると，出力アルファベットの生起確率を

$$q=(q_1,q_2,\cdots,q_s)$$

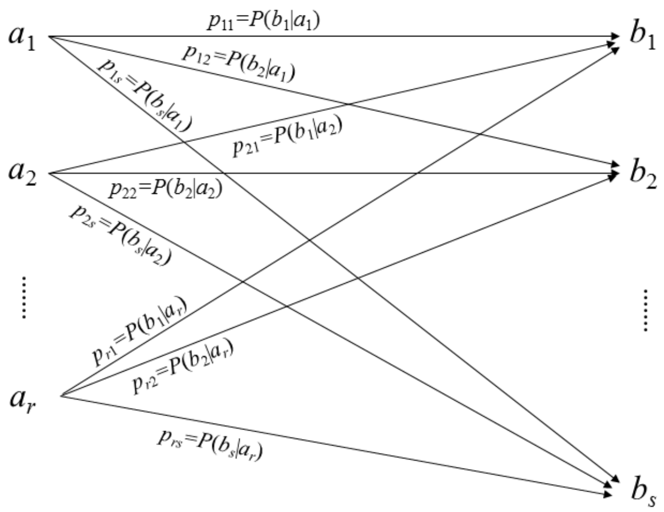

図 4.　通信路線図

とすると，

$$
\begin{aligned}
q &= (q_1, q_2, \cdots, q_s) \\
&= pT \\
&= (p_1, p_2, \cdots, p_r)\begin{bmatrix} p_{11} & \cdots & p_{1s} \\ \vdots & \ddots & \vdots \\ p_{r1} & \cdots & p_{rs} \end{bmatrix} \\
&= (p_1p_{11} + p_2p_{21} + \cdots + p_rp_{r1}, p_1p_{12} + p_2p_{22} + \cdots + p_rp_{r2}, \cdots, p_1p_{1s} + p_2p_{2s} + \cdots + p_rp_{rs})
\end{aligned}
$$

7.5 基本的な通信路

基本的な通信路として，2元対称通信路(Binary Symmetric Channel, BSC)と2元対称消失通信路(Binary Symmetric Erasure Channel, BSEC)がある．

(1) 2元対称通信路(Binary Symmetric Channel, BSC)

通信路行列：

$$T = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$ 　ここで，pは誤り率

通信路線図：図 5 の通り．

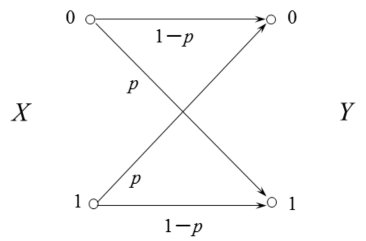

図 5.　BSC の通信路線図

(2) 2 元対称消失通信路(Binary Symmetric Erasure Channel, BSEC)

$0, 1$がXに化ける（＝消失する）ことがあるという点が BSC と異なる．

通信路行列：

$$T = \begin{bmatrix} 1 - p_{\mathrm{X}} - p & p_{\mathrm{X}} & p \\ p & p_{\mathrm{X}} & 1 - p_{\mathrm{X}} - p \end{bmatrix}$$

通信路線図：図 6 の通り．

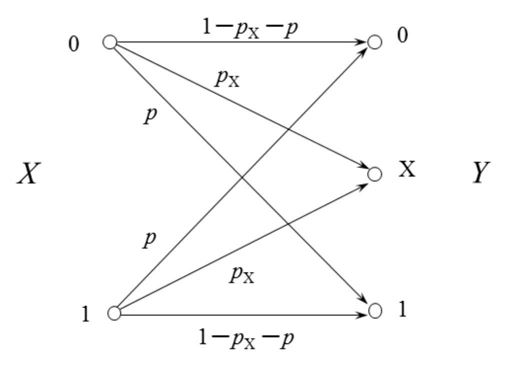

図 6．BSEC の通信路線図