

今後の講義予定

11/5 : 通常(第**7**回)

11/12 : 通常(第**8**回)

11/19 : 通常(第**9**回)

11/26 : 通常(第**10**回)

講義目次

- **7. 通信路**

- 7.1 通信路のモデル
- 7.2 離散通信路
- 7.3 記憶のない定常通信路
- 7.4 通信路行列
- 7.5 通信路線図
- 7.6 2元対称通信路(BSC)
- 7.7 2元非対称通信路
- 7.8 2元消失通信路(BEC)
- 7.9 完全非対称2元通信路
- 7.10 無雑音通信路
- 7.11 確定的通信路
- 7.12 一様通信路
- 7.13 通信路のモデル化について

講義目次

- **8.相互情報量と通信路容量**
 - 8.1 通信路に関する確率計算
 - 8.2 事前確率と事後確率
 - 8.3 事前エントロピーと事後エントロピー
 - 8.4 相互情報量の定義
 - 8.5 相互情報量 $I(A; B)$ の意味
 - 8.6 相互情報量 $I(A; B)$ の計算式
 - 8.7 $I(A; B)$ の計算
 - 8.8 $I(A; B)$ の計算例
 - 8.9 BSCの相互情報量: グラフ表現
 - 8.10 相互情報量の凸性
 - 8.11 相互情報量 $I(A; B)$ の性質
 - 8.12 3事象間の相互情報量

7. 通信路



- 7.1 通信路のモデル
- 7.2 離散通信路
- 7.3 記憶のない定常通信路
- 7.4 通信路行列
- 7.5 通信路線図
- 7.6 2元対称通信路(BSC)
- 7.7 2元非対称通信路
- 7.8 2元消失通信路(BEC)
- 7.9 完全非対称2元通信路
- 7.10 無雑音通信路
- 7.11 確定的通信路
- 7.12 一様通信路
- 7.13 通信路のモデル化について

7.1 通信路のモデル

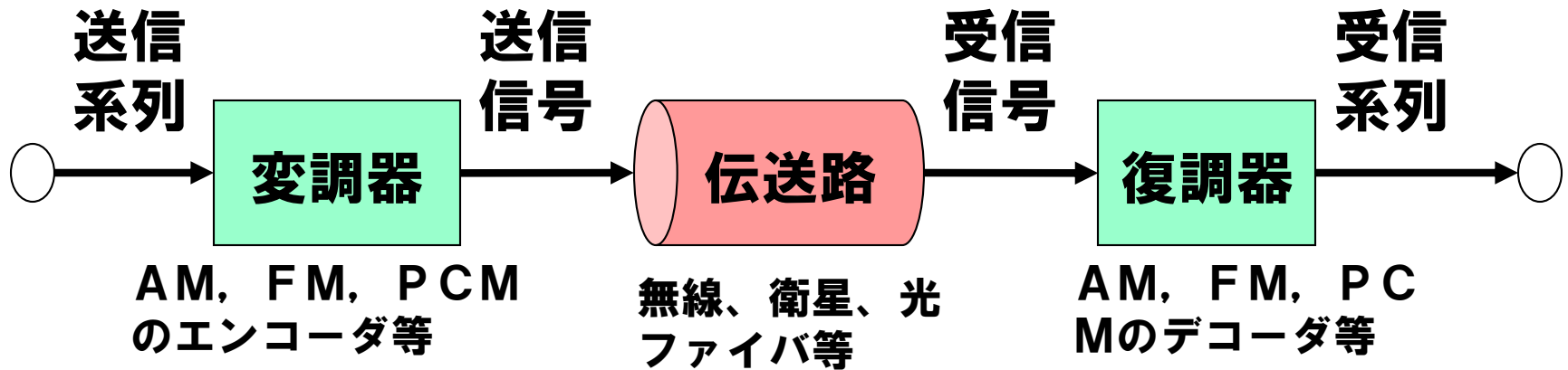


図1 通信路の一般的モデル

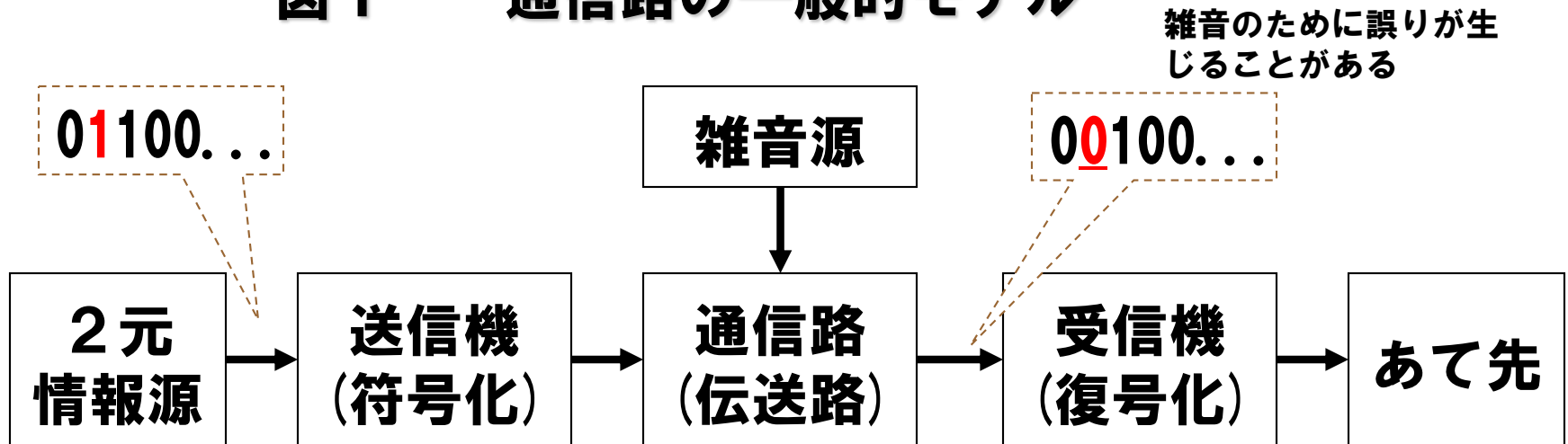
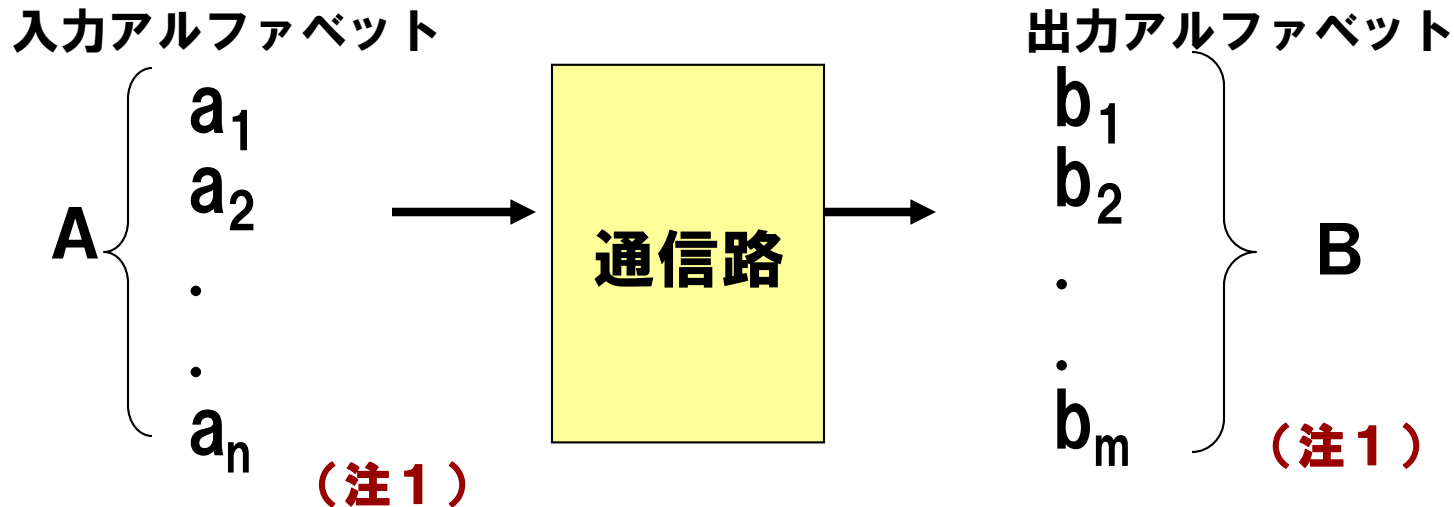


図2 2元データ通信路モデル

7.2 離散通信路

- 送信情報と受信情報がいずれもデジタル情報である通信路を、**離散通信路**(discrete channel)と呼ぶ



通信路の特性は、条件付確率

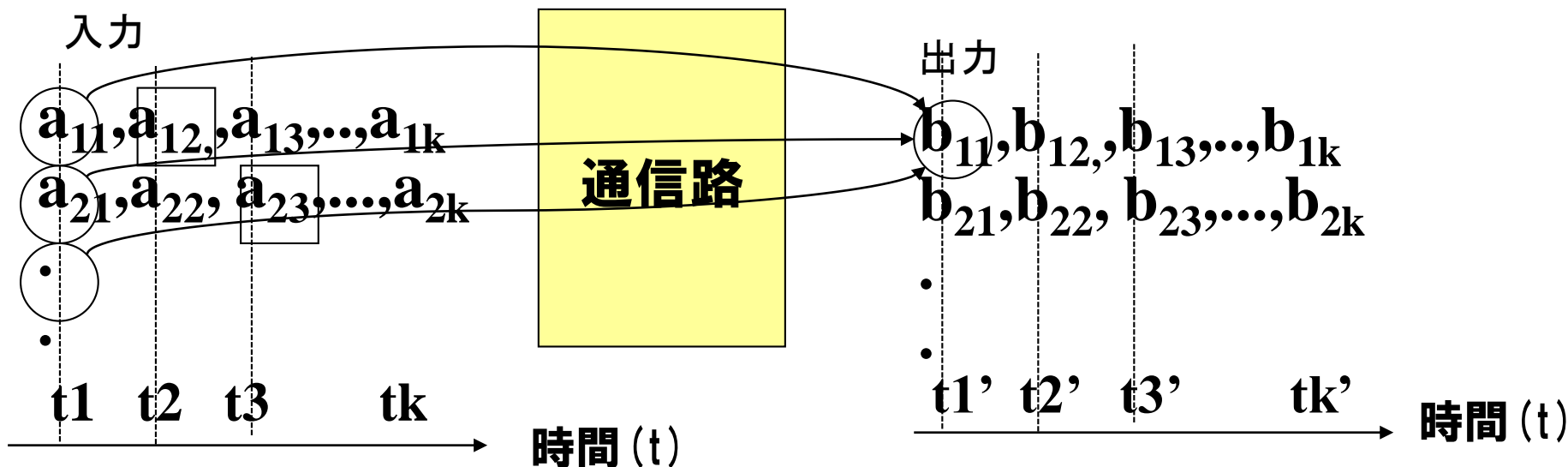
$P(b_j | a_i)$ により記述される (注2)

(注1) 雑音があるため、 a_i , b_j は1対1に対応せず、入力数(n)と出力数(m)は必ずしも等しくない

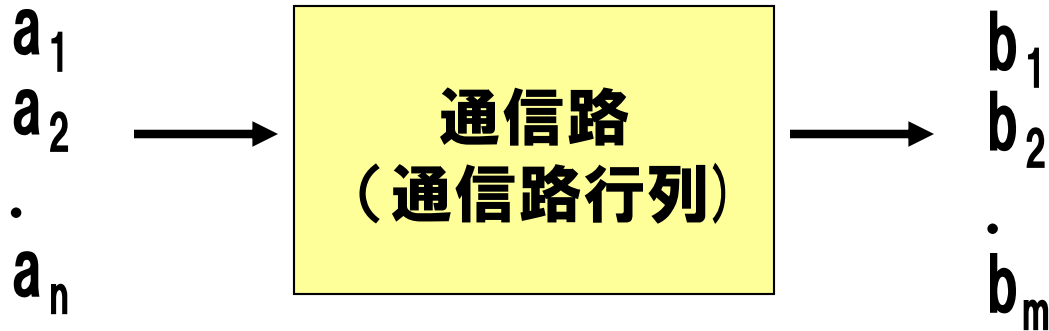
(注2) 記憶のない定常通信路の時に成り立つ

7.3 記憶のない定常通信路

- 各時点の出力の現れかたが、その時点の入力には関係するが、それ以外の時点の入力にも出力にも無関係(独立)となる時、**記憶のない通信路 (memoryless channel)**, という。
 - 時刻 t_1 の出力 b_{11} は、同じ時刻 t_1 の入力には関係するが、それ以外は無関係
- さらに、この通信路が定常である時(即ち、時間をずらしても統計的性質が変わらない時)、**記憶のない定常通信路 (memoryless stationary channel)**, という。
 - 例えば、時刻 t_1 の時の条件付き確率分布と、時刻 t_2, t_3, \dots の確率分布が同じである



7.4 離散通信路の表現法：通信路行列



列が出力シンボル

		b_1	b_2	b_3	\dots	b_m
行が入力シンボル	$P =$	P_{11}	P_{12}	P_{13}	\dots	P_{1m}
	a_1	P_{21}	P_{22}	P_{23}	\dots	P_{2m}
	a_2	\dots	\dots	\dots	\dots	\dots
	\dots	P_{n1}	P_{n2}	P_{n3}	\dots	P_{nm}
	a_n					

$$P_{ij} = P(b_j | a_i)$$

a_i を送信した時、 b_j を受信する確率

7.4 通信路行列の性質(1)

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \dots & P_{1m} \\ P_{21} & P_{22} & P_{23} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & P_{n3} & \dots & P_{nm} \end{pmatrix} \quad \text{総和} = 1$$

- 各行の総和は 1 になる。即ち、

$$\begin{aligned} \sum_j P_{ij} &= \\ P_{i1} + P_{i2} + \dots + P_{im} &= P(b_1 | a_i) + P(b_2 | a_i) + \dots + P(b_m | a_i) \\ &= 1 \end{aligned}$$

これはある a_i を送った時、 b_j のいずれかが必ず受信されるため。

7.4 通信路行列(2)

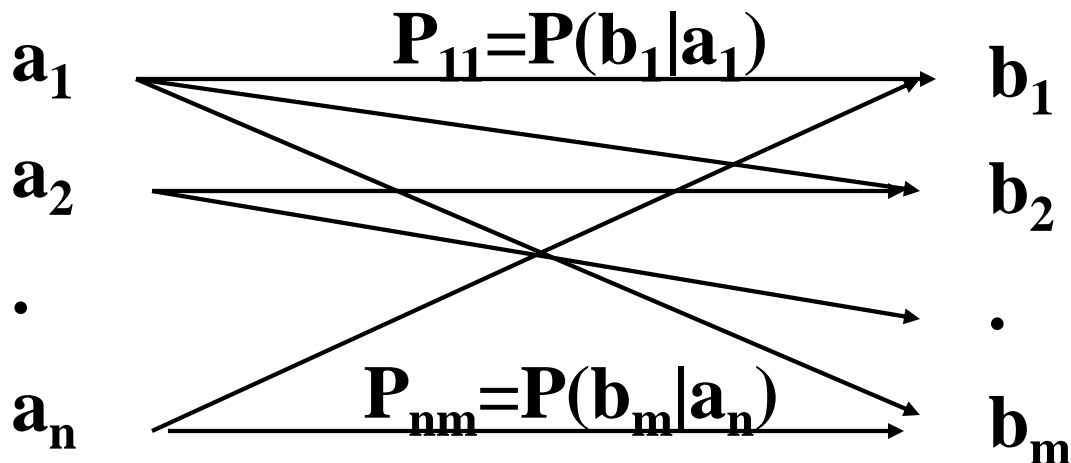
- 送信シンボル(a_i)の生起確率 $P(a_i)$ と通信路行列 P_{ij} とから、受信シンボル(b_j)の生起確率 $P(b_j)$ が計算できる

行列 P^t ← t: 転置行列

$$\begin{pmatrix} P(b_1) \\ P(b_2) \\ \cdot \\ \cdot \\ P(b_m) \end{pmatrix} = \begin{pmatrix} P_{11} & P_{21} & P_{31} & \cdot & \cdot & \cdot & P_{n1} \\ P_{12} & P_{22} & P_{32} & \cdot & \cdot & \cdot & P_{n2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ P_{1m} & P_{2m} & P_{3m} & \cdot & \cdot & \cdot & P_{nm} \end{pmatrix} \begin{pmatrix} P(a_1) \\ P(a_2) \\ \cdot \\ \cdot \\ P(a_n) \end{pmatrix}$$

$$\begin{aligned} P(b_1) &= P_{11}P(a_1) + P_{21}P(a_2) + \dots + P_{n1}P(a_n) \\ &= P(b_1|a_1)P(a_1) + P(b_1|a_2)P(a_2) + \dots + P(b_1|a_n)P(a_n) \end{aligned}$$

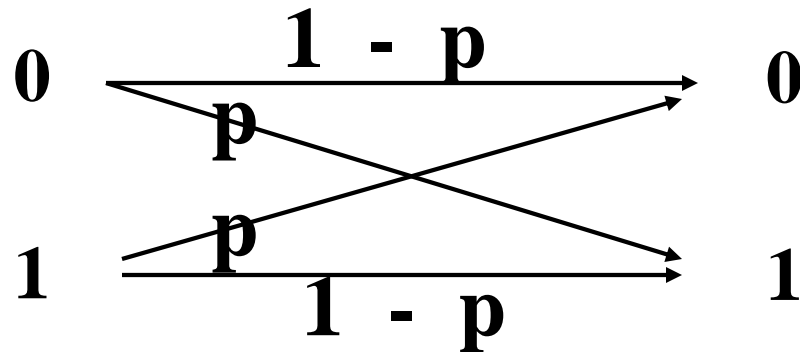
7.5 離散通信路の表現法：通信路線図



- 通信路を図で表したものを**通信路線図**と言う。
 - 送信シンボル a_i と受信シンボル b_j を矢印で結ぶ
 - 矢印上に、条件付き確率を書く
- => 各送信シンボルから出る矢印に付く確率の総和は全て1になる
($\sum_j P_{ij} = P_{i1} + P_{i2} + \dots + P_{im} = 1$)

添え字(i)について総和

7.6 2元対称通信路(BSC)



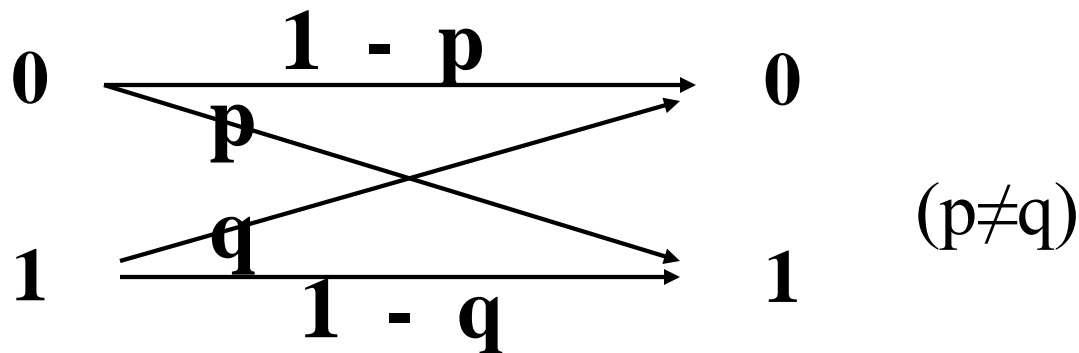
$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

■ BSCとは

- 0を送ったとき1が受信される確率 (p) と, 1を送った時0が受信される確率 (p) が等しい通信路

BSC: Binary Symmetric Channel

7.7 2元非対称通信路(一般の2元通信路)



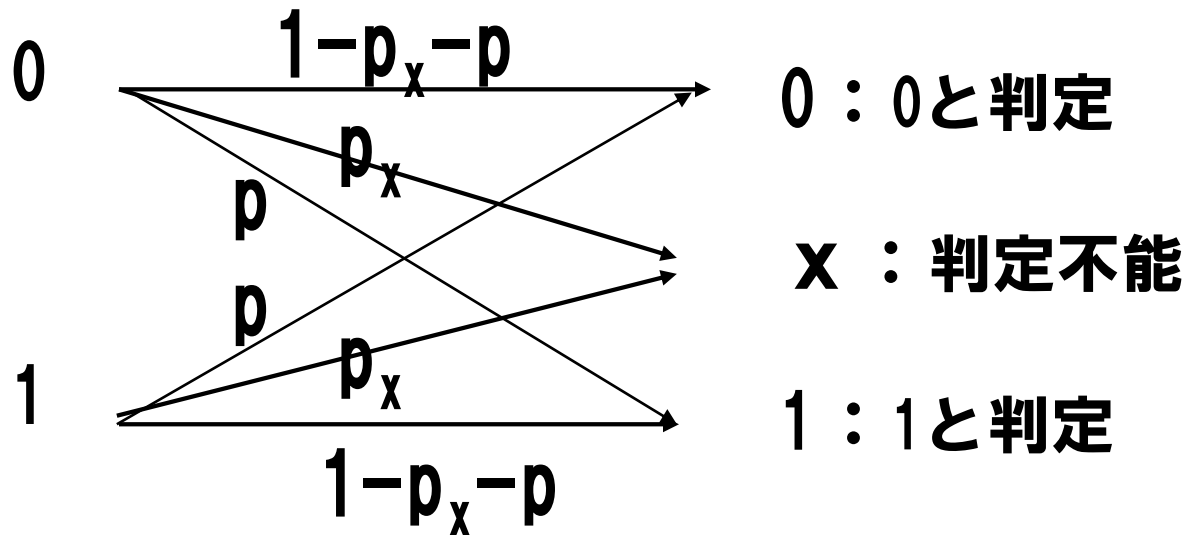
$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

- 一般の2元通信路(2元非対称通信路)とは
 - 0を送ったとき1が受信される確率 (p) と, 1を送った時0が受信される確率 (q) が異なる通信路
 - (例) $0 \rightarrow 1$ の誤り率 p が 0.2, $1 \rightarrow 0$ の誤り率 q が 0.6

7.8 2元消失通信路 (BEC) 一般形

■ BEC(一般形)とは

- 受信側で無理に判定を下さず、送信シンボル a_i が何か分からない状態 (x) を設けた通信路。

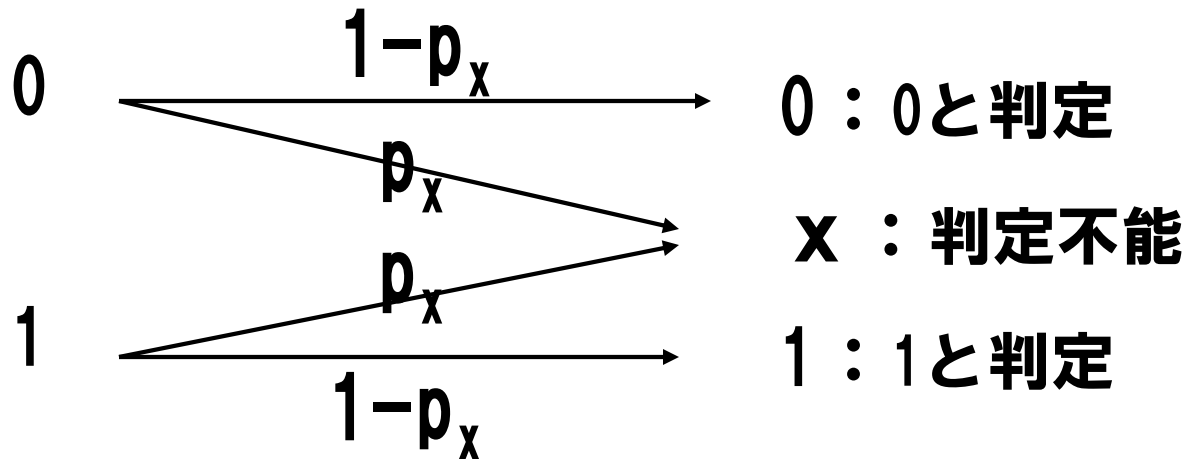


$$P = \begin{pmatrix} 1 - p_x - p & p_x & p \\ p & p_x & 1 - p_x - p \end{pmatrix}$$

7.8 2元消失通信路 (BEC)

■ BEC(通常形)とは

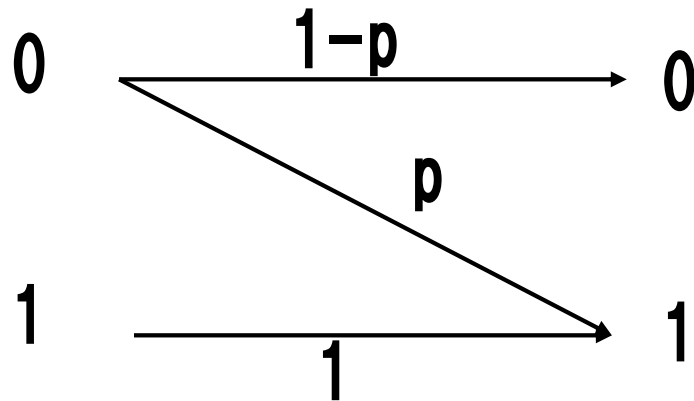
- 一般形BECの中で、0が1に誤ることも、1が0に誤ることもないもの (前ページ例で $p=0$ の場合)



$$P = \begin{pmatrix} 1-p_x & p_x & 0 \\ 0 & p_x & 1-p_x \end{pmatrix}$$

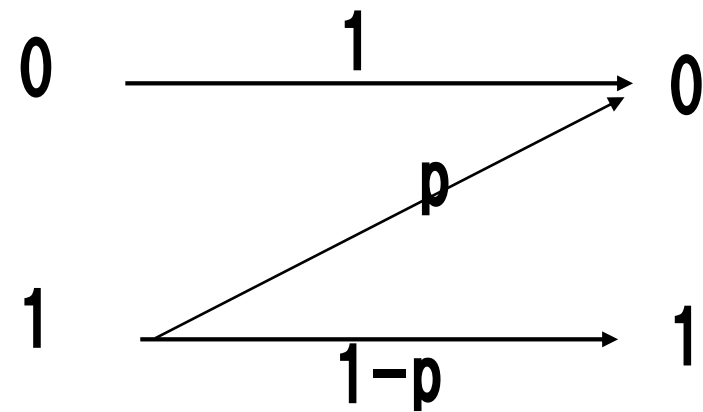
7.9 完全非対称2元通信路

- 0には雑音が加わって1になることがある。しかし、1は0にはならない(タイプ1の場合)
- 1には雑音が加わって0になることがある。しかし、0は1にはならない(タイプ2の場合)



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

(タイプ1)

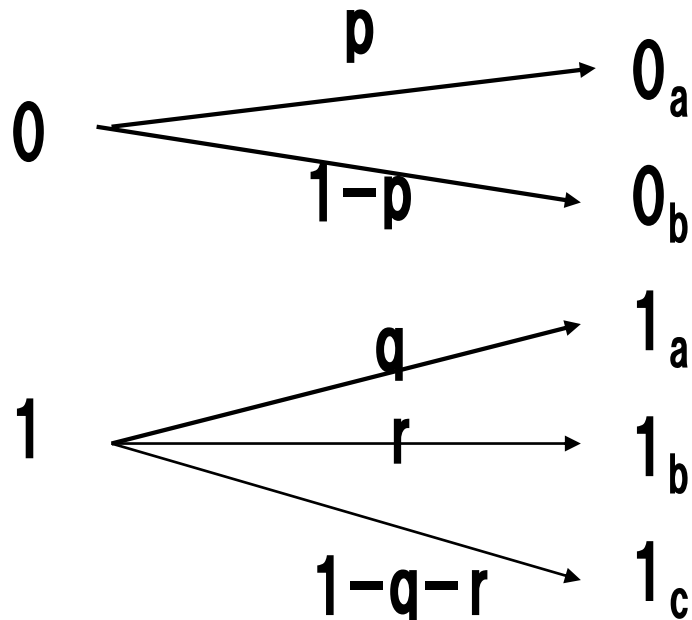


$$P = \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}$$

(タイプ2)

7.10 無雑音通信路

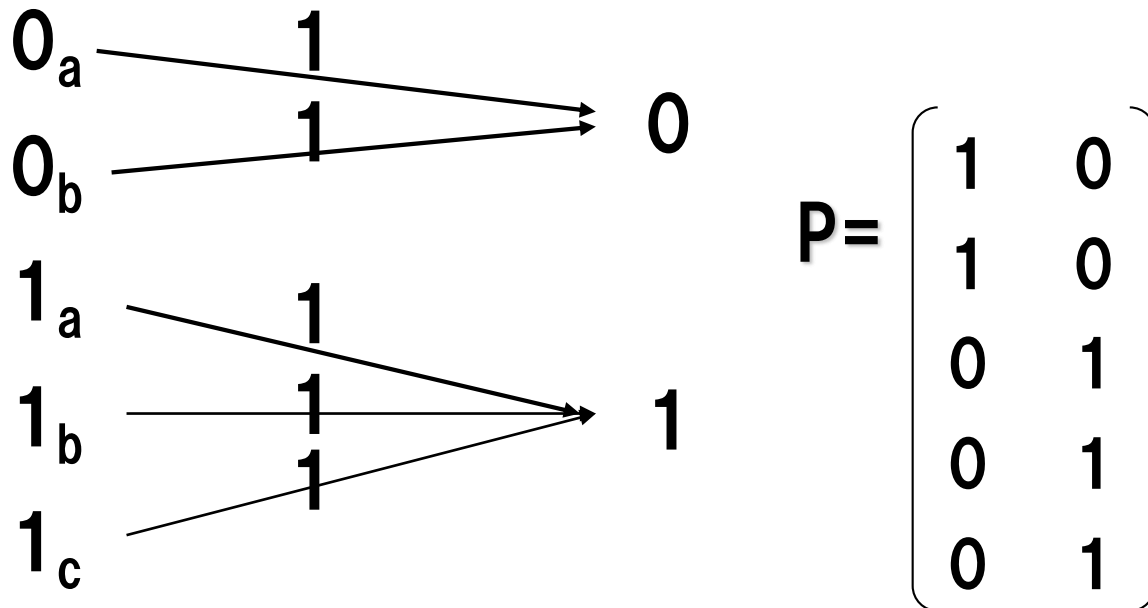
- 0には雑音が加わって $0_a, 0_b$ になることがある。しかし、 $1_a \sim 1_c$ にはならない
- 1には雑音が加わって $1_a \sim 1_c$ になることがある。しかし $0_a, 0_b$ にはならない。
- つまり、どちらの送信シンボルも受信シンボル範囲が重ならず、雑音の影響を受けない



$$P = \begin{pmatrix} p & 1-p & 0 & 0 & 0 \\ 0 & 0 & q & r & 1-q-r \end{pmatrix}$$

7.11 確定的通信路

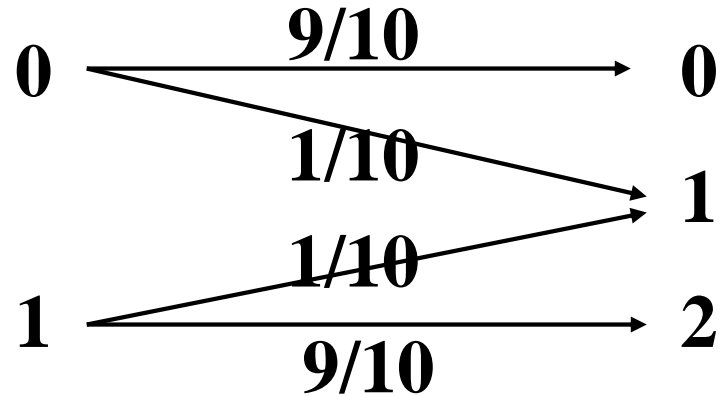
- $0_a, 0_b$ ともに0にしかない
- $1_a \sim 1_c$ ともに1にしかない
- つまり, どの送信シンボルも受信シンボルが一意に確定する



7.12 一様通信路

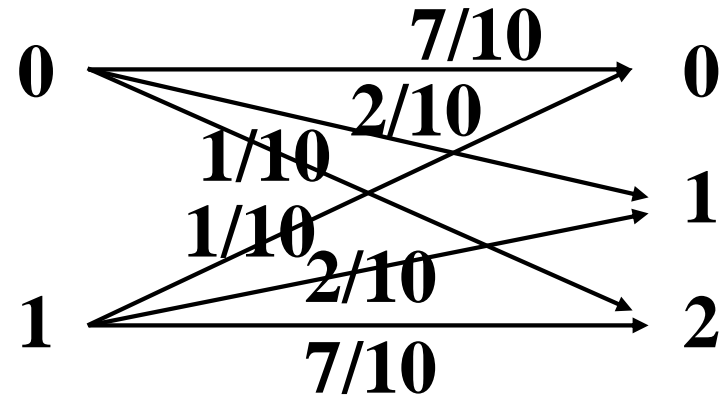
- 通信路の中である種の対象性あるいは均一性をもっているものは、相互情報量を容易に計算できる、特に次の一様通信路と呼ばれるものは重要。
- 「入力から見て一様」な通信路
 - 通信路行列の各行が全く等しい要素の組をもち、1つの行を適当に置き換えることにより他の行が得られる。
 - **どの入力要素も出力への影響が同じ性質**(入力が均質である)。
- 「出力から見て一様」な通信路
 - 通信路行列の各列が全く等しい要素の組をもち、1つの列を適当に置き換えることにより他の列が得られる。
 - **どの出力要素も入力からの影響が同じ**(出力が均質である)。
- 「2重に一様」な通信路
 - 入力から見ても、出力から見ても一様な通信路

7.12 入力から見て一様通信路



$$\begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 9/10 & 1/10 & 0 \\ 0 & 1/10 & 9/10 \end{pmatrix} \end{matrix}$$

(a) BEC

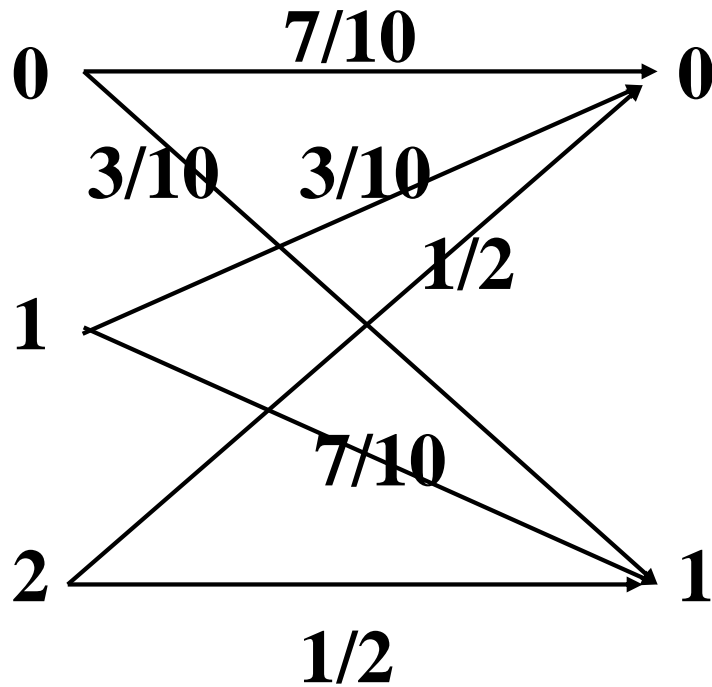


$$\begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 7/10 & 2/10 & 1/10 \\ 1/10 & 2/10 & 7/10 \end{pmatrix} \end{matrix}$$

(b)

- 各行が等しい要素の組をもつ
- ある行を列入替えすると他の行と一致

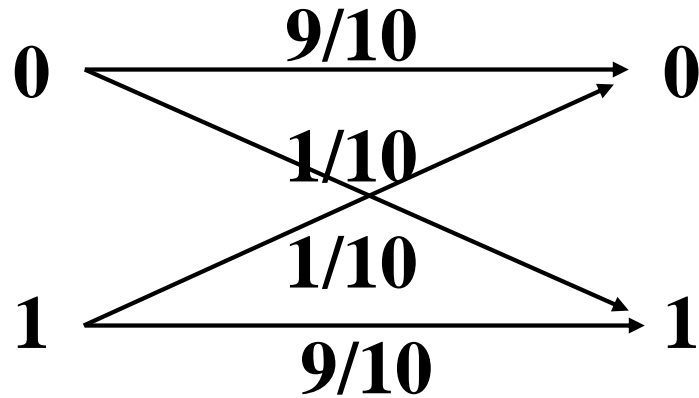
7.12 出力から見て一様通信路



$$\begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 0 & 1 \\ 7/10 & 3/10 \\ 3/10 & 7/10 \\ 1/2 & 1/2 \end{pmatrix}$$

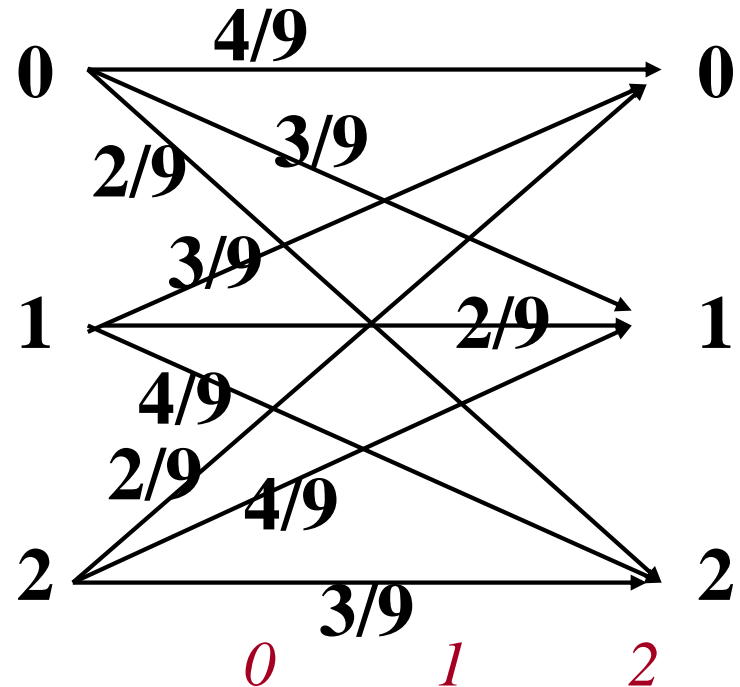
- 各列が等しい要素の組をもつ
- ある列を行入替えすると他の列と一致

7.12 一様通信路の例: 2重に一様



$$\begin{matrix}
 & \begin{matrix} 0 & 1 \end{matrix} \\
 \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 9/10 & 1/10 \\ 1/10 & 9/10 \end{bmatrix}
 \end{matrix}$$

(a) BSC



$$\begin{matrix}
 & \begin{matrix} 0 & 1 & 2 \end{matrix} \\
 \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{bmatrix} 4/9 & 3/9 & 2/9 \\ 3/9 & 2/9 & 4/9 \\ 2/9 & 4/9 & 3/9 \end{bmatrix}
 \end{matrix}$$

(b)

7.13 通信路のモデル化について

- 情報理論では、**通信路の統計的性質**が与えられると仮定しているが、現実の通信路ではこれは例外的である。
 - さらに、**情報源の統計的性質も与えられない**場合が多いし、情報源と通信路の分離があいまいな場合もある。
- 現実の通信路のモデル化は、**経験的あるいは実験的に決められる**場合が多い。
 - 例えばある一定期間、通信路を観測し、その期間に起きた誤り発生率をこの通信路のランダム誤り率としてモデル化する。しかしこの観測期間が**適当かどうか**わからない（厳密には統計的手法で確かめるべきである）。
 - **バースト誤りを発生する通信路ではさらに複雑になる。**
- **通信路のモデル化は、常に研究テーマである**

8. 相互情報量と通信路容量

- 8.1 通信路に関する確率計算
- 8.2 事前確率と事後確率
- 8.3 事前エントロピーと事後エントロピー
- 8.4 相互情報量の定義
- 8.5 相互情報量 $I(A; B)$ の意味
- 8.6 相互情報量 $I(A; B)$ の計算式
- 8.7 $I(A; B)$ の計算
- 8.8 $I(A; B)$ の計算例
- 8.9 BSCの相互情報量: グラフ表現
- 8.10 相互情報量の凸性
- 8.11 相互情報量 $I(A; B)$ の性質
- 8.12 3事象間の相互情報量

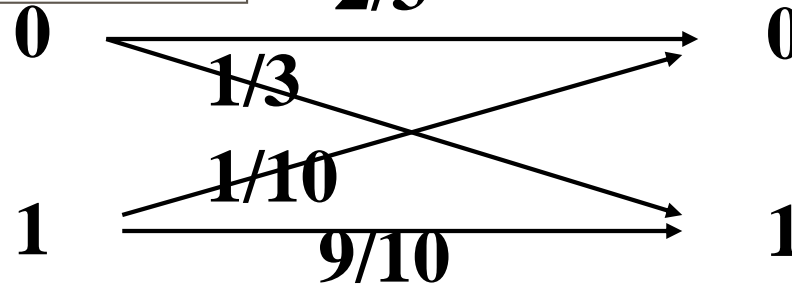
8.1 通信路に関する確率計算 (1)

情報源の確率

$$P(0)=3/4$$

$$P(1)=1/4$$

通信路の確率



0が1に誤る確率は比較的高いが、1が0に誤る確率は低い通信路。BSCではない。

$$P = \begin{pmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{pmatrix} \quad P^t = \begin{pmatrix} 2/3 & 1/10 \\ 1/3 & 9/10 \end{pmatrix}$$

$$P(b=0) = (2/3) * (3/4) + (1/10) * (1/4) = 21/40$$

$$P(b=1) = (1/3) * (3/4) + (9/10) * (1/4) = 19/40$$

8.1 通信路に関する確率計算 (2)

- ・ 受信後の送信シンボルの確率を計算するため、次の公式(ベイズの定理)を利用する。

$$\begin{aligned}P(a_i|b_j) &= P(a_i, b_j) / P(b_j) = P(b_j|a_i) P(a_i) / P(b_j) \\ &= P(b_j|a_i) P(a_i) / \sum_i P(b_j|a_i) P(a_i)\end{aligned}$$

- ・ 計算例

$$P(a=0 | b=0) = (2/3) * (3/4) / (21/40) = 20/21$$

$$P(a=1 | b=0) = (1/10) * (1/4) / (21/40) = 1/21$$

$$P(a=0 | b=1) = (1/3) * (3/4) / (19/40) = 10/19$$

$$P(a=1 | b=1) = (9/10) * (1/4) / (19/40) = 9/19$$

- ・ 送信シンボルと受信シンボルの結合確率

$$P(a=0, b=0) = P(b=0|a=0) * P(a=0) = 1/2$$

$$P(a=0, b=1) = P(b=1|a=0) * P(a=0) = 1/4 \text{ など}$$

8.2 事前確率と事後確率

- 入力シンボル a_i の確率 $P(a_i)$ を**事前確率**、受信シンボル b_j を観測した後の入力シンボルの確率 $P(a_i|b_j)$ を**事後確率**と呼ぶ
 - 前の例では、事前確率は、 $P(a=0) = 3/4 = 0.75$,
 - 事後確率は、
 $P(a=0|b=0) = 20/21 = 0.952$,
 $P(a=0|b=1) = 10/19 = 0.526$
 - (意味)最初 $a=0$ の起きる確率が0.75であったものが、
 $b=0$ (同じシンボル)を受信すると、 $a=0$ の起きた確率が高くなるが、
 $b=1$ (違うシンボル)を受信すると、 $a=0$ の起きた確率が低くなる

8.3 事前エントロピーと事後エントロピー

- ・ 情報源AのエントロピーをAの事前エントロピーと呼ぶ

$$H(A) = \sum_i P(a_i) \log(1/P(a_i)) = -\sum_i P(a_i) \log P(a_i)$$

- ・ 出力シンボル b_j を受信後のAの事後エントロピー

$$H(A|b_j) = -\sum_i P(a_i | b_j) \log P(a_i | b_j)$$

(ある受信シンボルを知った時、入力シンボルのもつエントロピー)

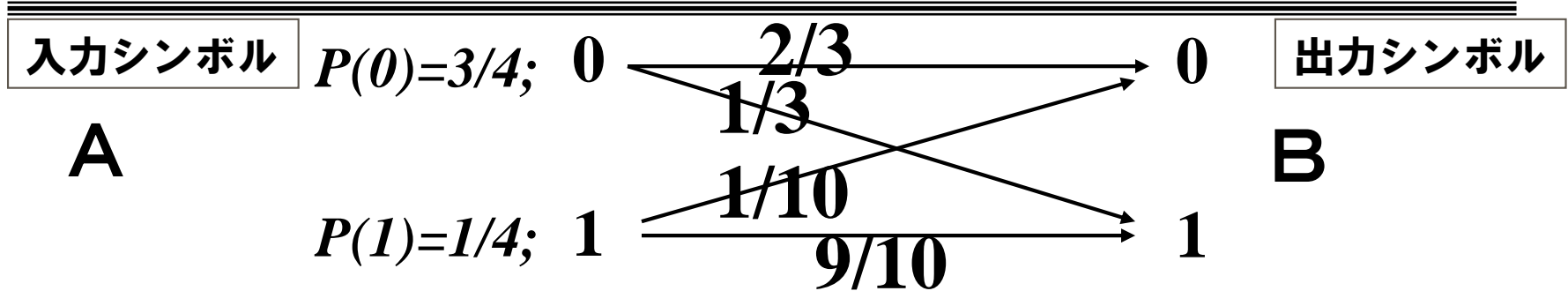
- ・ 全ての b_j の生起確率について平均した以下のものを平均事後エントロピーと呼ぶ。これは『AのBに関するあいまい度』とも言う。

$$H(A|B) = \sum_j P(b_j) H(A|b_j) = -\sum_j \sum_i P(b_j) P(a_i | b_j) \log P(a_i | b_j) = -\sum_i \sum_j P(a_i, b_j) \log P(a_i | b_j)$$

- ・ 同様に、『BのAに関するあいまい度』は下記で定義される。これを散布度とも言う。

$$H(B|A) = -\sum_j \sum_i P(a_i) P(b_j | a_i) \log P(b_j | a_i)$$

8.3 事前、事後エントロピーの計算



- **Aの事前エントロピーは、**

$$H(A) = (3/4) * \log(4/3) + (1/4) * \log 4 = \underline{0.811} \text{ (bit)}$$

- **出力b=0を観測後のAの事後エントロピーは、**

$$P(a=0|b=0) = 20/21, P(a=1|b=0) = 1/21 \text{ を用いて、}$$

$$H(A|0) = (20/21) \log(21/20) + (1/21) \log 21 = \underline{0.276}$$

- **出力b=1を観測後のAの事後エントロピーは、**

$$P(a=0|b=1) = 10/19, P(a=1|b=1) = 9/19 \text{ を用いて、}$$

$$H(A|1) = (10/19) \log(19/10) + (9/19) \log(19/9) = \underline{0.998}$$

- **平均事後エントロピー $H(A|B) = \underline{0.619}$**

8.4 相互情報量の定義

- 事前エントロピーと平均事後エントロピーの差を**相互情報量** $I(A;B)$ と定義する (注)

$$I(A;B) = H(A) - H(A|B)$$

$$= \sum_{A,B} P(a_i, b_j) \log (P(a_i|b_j) / P(a_i))$$

$$= \sum_{A,B} P(a_i, b_j) \log (P(a_i, b_j) / P(a_i) P(b_j))$$

- $A = \{0, 1\}$, $B = \{0, 1\}$ の場合、

$$I(A; B) = P(0,0) \log P(0,0)/P(a=0)P(b=0) + P(0,1) \log P(0,1)/P(a=0)P(b=1) + P(1,0) \log P(1,0)/P(a=1)P(b=0) + P(1,1) \log P(1,1)/P(a=1)P(b=1)$$

(注) $I(A;B)$ の「:」はセミコロンであり, 「,」コンマではない

8.5 相互情報量 $I(A; B)$ の意味

$H(A) - H(A|B)$ の意味・解釈:

- 通報 b_j を知ることで a_i に関する曖昧性が減少する割合を示す。
すなわち、「**情報源の情報量のうち、通信路により受信側に伝わる情報量(伝達情報量)**」を表す。
 - なお相互情報量は、通信路の入出力事象間に限らず、一般的に2つ以上の事象系の間のエントロピーの差として定義される。それら「**事象間に相関性がどれだけあるか**」を測る尺度でもある。
- 前例の通信路の相互情報量は、
$$I(A; B) = [H(A) = 0.811] - [H(A|B) = 0.619] = 0.192$$

Aの曖昧性0.811ビットが、受信したBを観測する(知る)ことで0.619ビットに減少した。したがって、これは0.192ビットだけ受信側に情報が伝送された、と解釈できる

8.6 相互情報量 $I(A; B)$ の計算式

- 事前エントロピーと平均事後エントロピーの差が **相互情報量 $I(A; B)$**

$$I(A; B) = H(A) - H(A|B) \quad [\text{定義}]$$

$$= -\sum_A P(a_i) \log P(a_i) + \sum_{A,B} P(a_i, b_j) \log P(a_i|b_j) \quad \text{定義の展開}$$

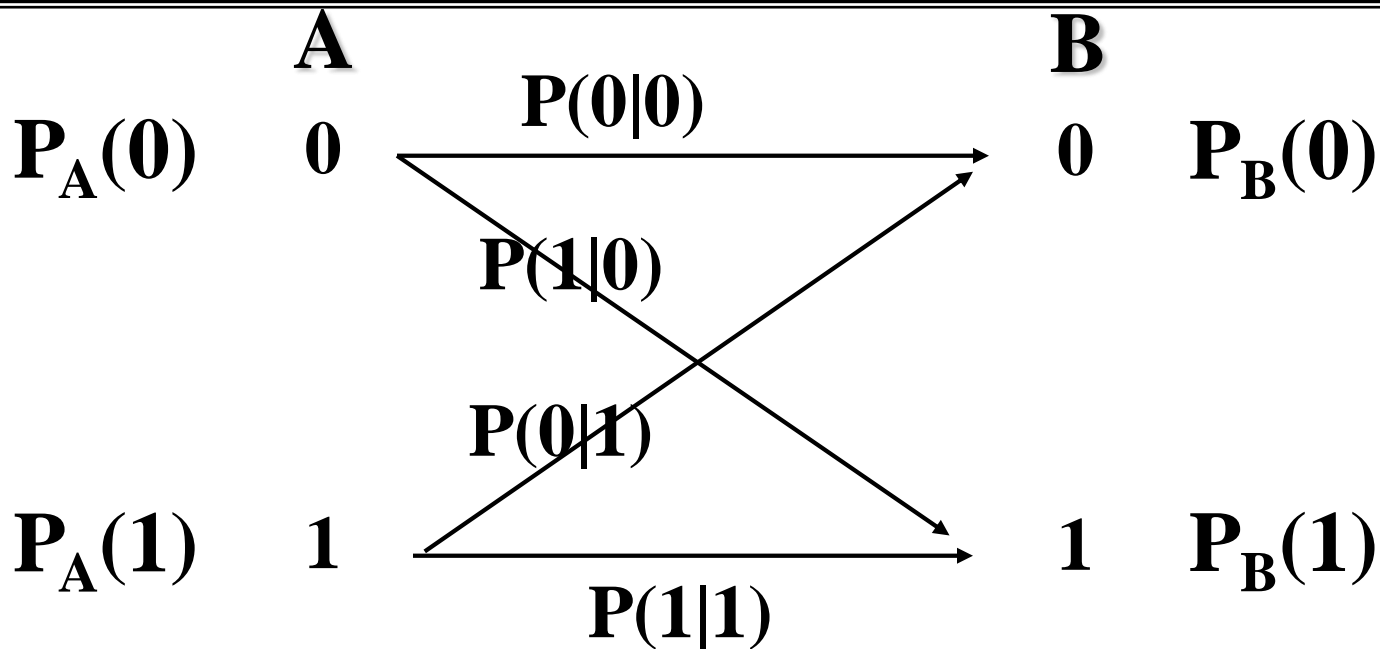
$$= \sum_{A,B} [P(a_i, b_j) \log P(a_i|b_j) - \log P(a_i)] \quad \textcircled{1}$$

$$= \sum_{A,B} P(a_i, b_j) \log [P(a_i|b_j) / P(a_i)] \quad \textcircled{2}$$

$$= \sum_{A,B} P(a_i, b_j) \log [P(a_i, b_j) / P(a_i) P(b_j)] \quad \textcircled{3}$$

- ① a_i の事後確率 $P(a_i|b_j)$ と同時確率 $P(a_i, b_j)$ が容易に求まるとき
- ② 各事象ごとの相互情報量 $\log \{P(a_i|b_j) / P(a_i)\}$ が容易に求まるとき
- ③ 周辺確率 $P(a_i)$, $P(b_j)$ と同時確率 $P(a_i, b_j)$ が容易に求まるとき

8.7 I (A;B) の計算: 2元通信路の確率表現



通信路行列

各シンボルの通信路での遷移確率を行列で表わしたもの

$$P = \begin{pmatrix} P(0|0) & P(1|0) \\ P(0|1) & P(1|1) \end{pmatrix}$$

(簡略表記の説明)

$P_A(0)$ は $P(a=0)$, $P_A(1)$ は $P(a=1)$
 $P_B(0)$ は $P(b=0)$, $P_B(1)$ は $P(b=1)$

$P(0|0)$ は, $P(b=0|a=0)$,
 $P(1|0)$ は, $P(b=1|a=0)$

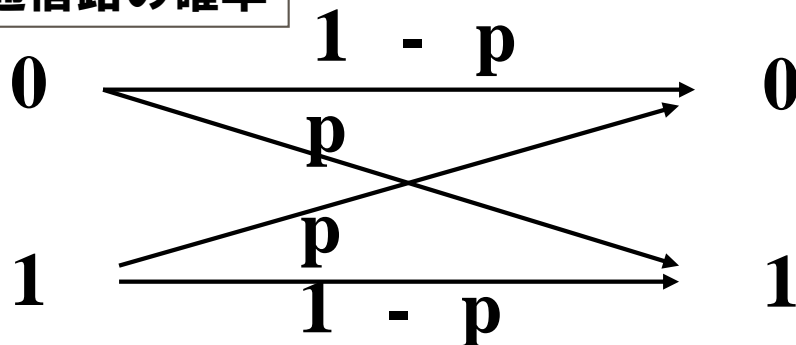
8.8 I (A;B) の計算例：BSCの場合

情報源の確率

$$P(0)=w$$

$$P(1)=1-w$$

通信路の確率



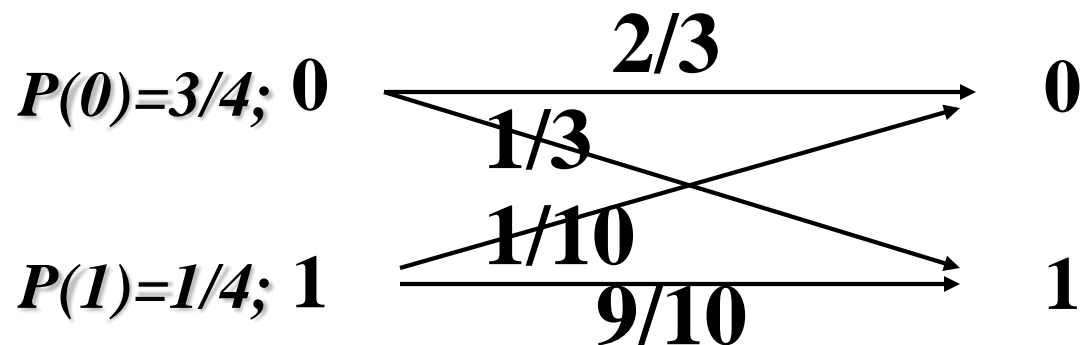
$$P(b=0) = w(1-p) + (1-w)p = 1 - [wp + (1-w)(1-p)]$$

$P(b=1) = wp + (1-w)(1-p)$ であるので、 $H(B) = -P(b=0) \log P(b=0) - P(b=1) \log P(b=1) = H(wp + (1-w)(1-p))$ となる。ここで $H(x)$ は x のエントロピー関数。一方、 $H(B|A) = H(p)$ となるので、結局

$I(A;B) = H(wp + (1-w)(1-p)) - H(p)$ 。 w と p の2変数関数。

エントロピー関数は、 $H(x) = -x \log x - (1-x) \log (1-x)$ の形。

8.8 I (A;B) の計算：通信路の例1



$$\mathbf{P} = \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix} \quad \mathbf{P}^t = \begin{bmatrix} 2/3 & 1/10 \\ 1/3 & 9/10 \end{bmatrix}$$

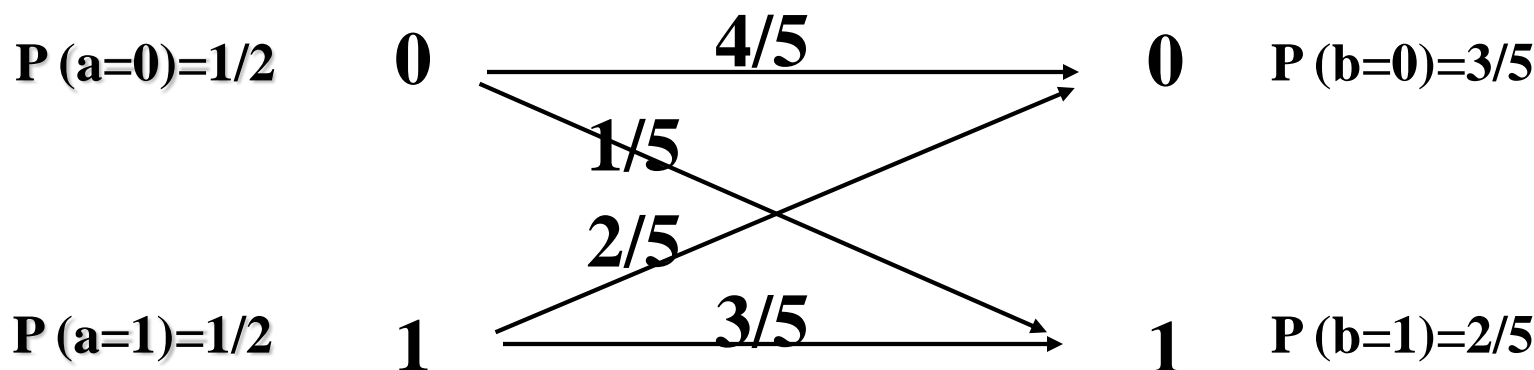
$$P(b=0) = (3/4) * (2/3) + (1/4) * (1/10) = 21/40$$

$$P(b=1) = (1/3) * (3/4) + (9/10) * (1/4) = 19/40$$

8.8 I (A;B) の計算: H (A) , H (A|B)

- $H(A) = -P(0) \log P(0) - P(1) \log P(1)$
 $= -3/4 * \log(3/4) - 1/4 * \log(1/4) = \underline{0.811}$
- $H(A|0) = -P(0|0) \log P(0|0) - P(1|0) \log P(1|0)$
 $= -20/21 * \log(20/21) - 1/21 * \log(1/21) = 0.276$
- $H(A|1) = -P(0|1) \log P(0|1) - P(1|1) \log P(1|1)$
 $= -10/19 * \log(10/19) - 9/19 * \log(9/19) = 0.998$
- $H(A|B) = H(A|0) * P(b=0) + H(A|1) * P(b=1)$
 $= 0.276 * (21/40) + 0.998 * (19/40) = \underline{0.619}$
- $I(A;B) = H(A) - H(A|B) = 0.811 - 0.619 = 0.192$

8.8 I (A;B) の計算：通信路の例2



$$P = \begin{bmatrix} 4/5 & 1/5 \\ 2/5 & 3/5 \end{bmatrix} \quad P^t = \begin{bmatrix} 4/5 & 2/5 \\ 1/5 & 3/5 \end{bmatrix} \cdot \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

(注) この通信路は対称でないので BSC
ではない

周辺確率(=各々の受信シンボルの確率)を計算すると:

$$P(b=0) = (1/2) * (4/5) + (1/2) * (2/5) = 3/5$$

b=0を受信するのは、a=0が送られて誤りが起きない場合、とa=1が送られ誤りが起きる場合、のどちらかであるので、両者の論理和となる。

$$P(b=1) = (1/2) * (1/5) + (1/2) * (3/5) = 2/5$$

8.8 I (A;B) の計算：事象毎の相互情報量 I (a_i;b_j)

- $I (a=1; b=1) = \log [P (a=1|b=1) / P (a=1)]$
 $= \log (3/4) / (1/2) = \log 3/2 = 0.59$
- $I (a=1; b=0) = \log [P (a=1|b=0) / P (a=1)]$
 $= \log (1/3) / (1/2) = \log 2/3 = -0.59$
- $I (a=0; b=1) = \log [P (a=0|b=1) / P (a=0)]$
 $= \log (1/4) / (1/2) = \log 1/2 = -1$
- $I (a=0; b=0) = \log [P (a=0|b=0) / P (a=0)]$
 $= \log (2/3) / (1/2) = \log 4/3 = 0.41$
- $I (A;B) = P (a=1, b=1) * I (a=1; b=1) + P (a=1, b=0) * I (a=1; b=0)$
 $+ P (a=0, b=1) * I (a=0; b=1) + P (a=0, b=0) * I (a=0; b=0)$
 $= (3/10) * 0.59 + (1/5) * (-0.59) + (1/10) * (-1) + (2/5) * 0.41$
 $= 0.12$

ある事象間の
相互情報量が
負になること
がある

8.8 I (A;B) の計算: H (A) とH (A|B) の差分

- $H(A) = -P(a=0)\log P(a=0) - P(a=1)\log P(a=1)$
 $= -1/2 * \log (1/2) - 1/2 * \log (1/2) = 1$

[1]

受信シンボル**b=0**、を固定して、**全てのa (a=0, 1) についてのエントロピー**

- $H(A|0) \dots [1]$

$$= -P(a=0|b=0)\log P(a=0|b=0) - P(a=1|b=0)\log P(a=1|b=0)$$
$$= -2/3 * \log (2/3) - 1/3 * \log (1/3) = 0.92$$

[2]

受信シンボル**b=1**、を固定して、**全てのa (a=0, 1) についてのエントロピー**

- $H(A|1) \dots [2]$

$$= -P(a=0|b=1)\log P(a=0|b=1) - P(a=1|b=1)\log P(a=1|b=1)$$
$$= -1/4 * \log (1/4) - 3/4 * \log (3/4) = 0.82$$

- $H(A|B) \dots [3]$

$$= H(A|0) * P(b=0) + H(A|1) * P(b=1)$$
$$= 0.92 * (3/5) + 0.82 * (2/5) = 0.88$$

[3]

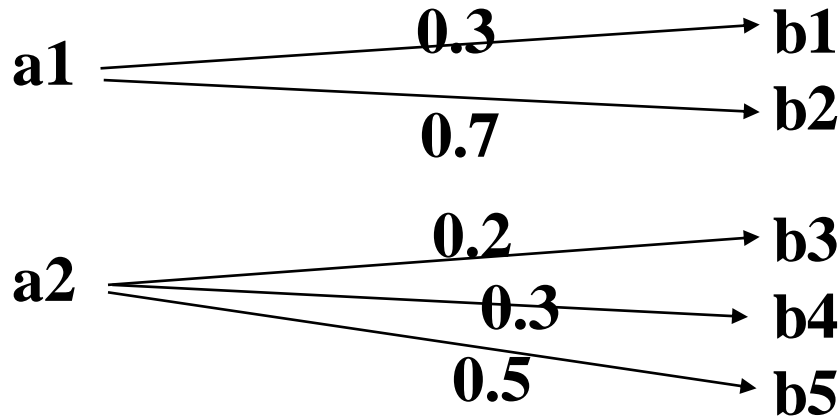
上記2つの事象毎条件付エントロピーを、**受信シンボルb (b=0,1) の生起確率により平均**をとったものの

- $I(A;B) = H(A) - H(A|B) = 1 - 0.88 = 0.12 \dots [4]$

[4]

I (A;B) は、Bを受信前のAの曖昧性 H (A) から、Bを受信後のAの曖昧性 H (A|B) を引いたもの

8.8 相互情報量の計算：無雑音通信路

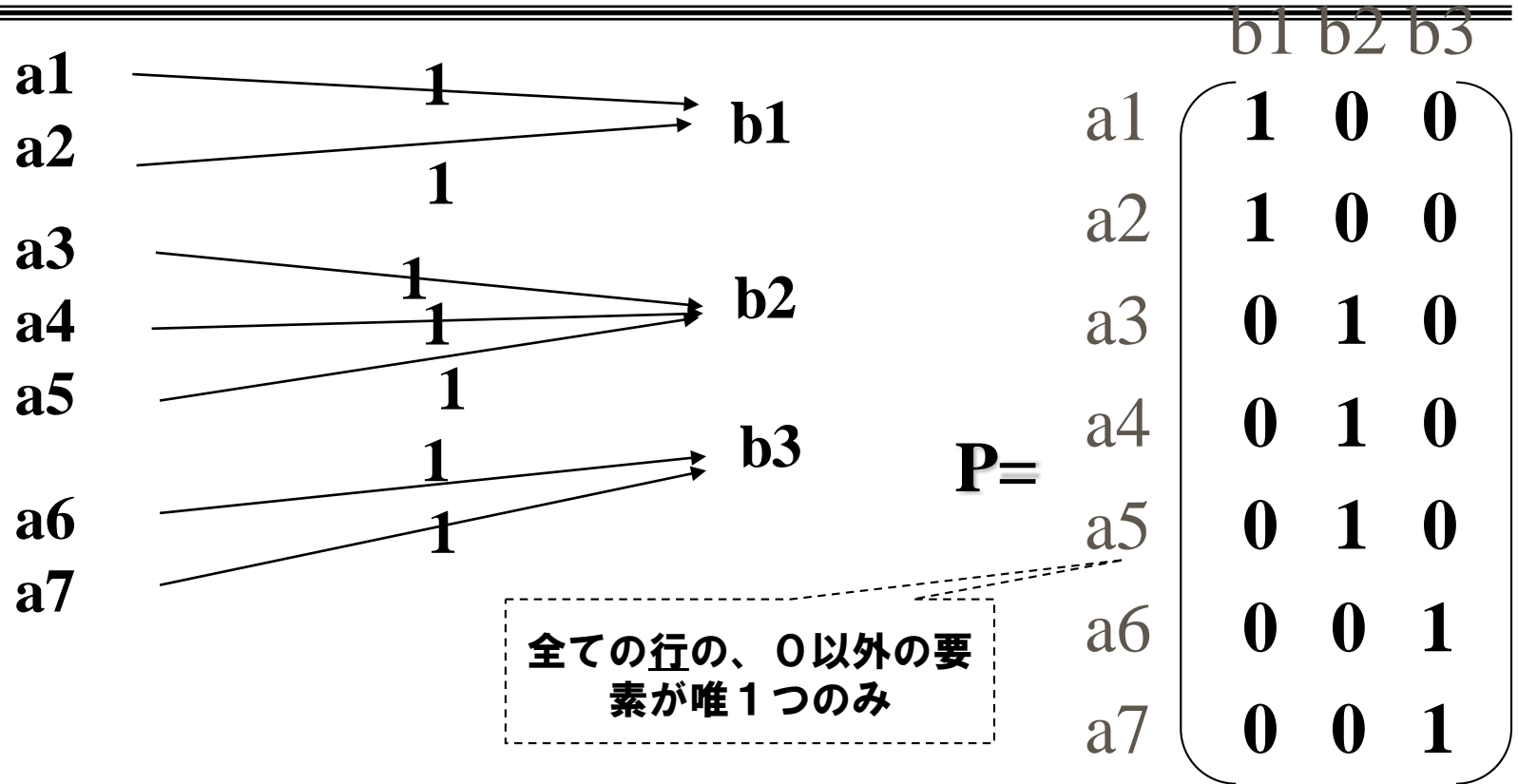


$$\mathbf{P} = \begin{matrix} & \begin{matrix} b1 & b2 & b3 & b4 & b5 \end{matrix} \\ \begin{matrix} a1 \\ a2 \end{matrix} & \begin{pmatrix} 0.3 & 0.7 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} \end{matrix}$$

全ての列の、0以外の要素が唯1つのみ

- $H(A|B) = 0$
 - 受信シンボル b_j を受信すると、送信シンボル a_i が一意に定まる、即ちAに関する曖昧性が0になる
 - 即ち、Bの事象を知るとAの事象が確定する
- $I(A; B) = H(A) - H(A|B) = H(A)$

8.8 相互情報量の計算：確定的通信路



- $H(B|A) = 0$

送信シンボル a_i が決まれば、受信シンボル b_j が一意に定まる、即ちBに関する曖昧性が0になる

- 即ち、Aの事象を知るとBの事象が確定する

- $I(A; B) = H(B) - H(B|A) = H(B)$

8.8 一様通信路の相互情報量

- 2重に一様な通信路の相互情報量は、次の簡単な式により求められる：

- $$I(A;B) = H(B) - H(B | A)$$
$$= H(B) - \sum_j P(b_j|a_i) \log(1/P(b_j|a_i))$$

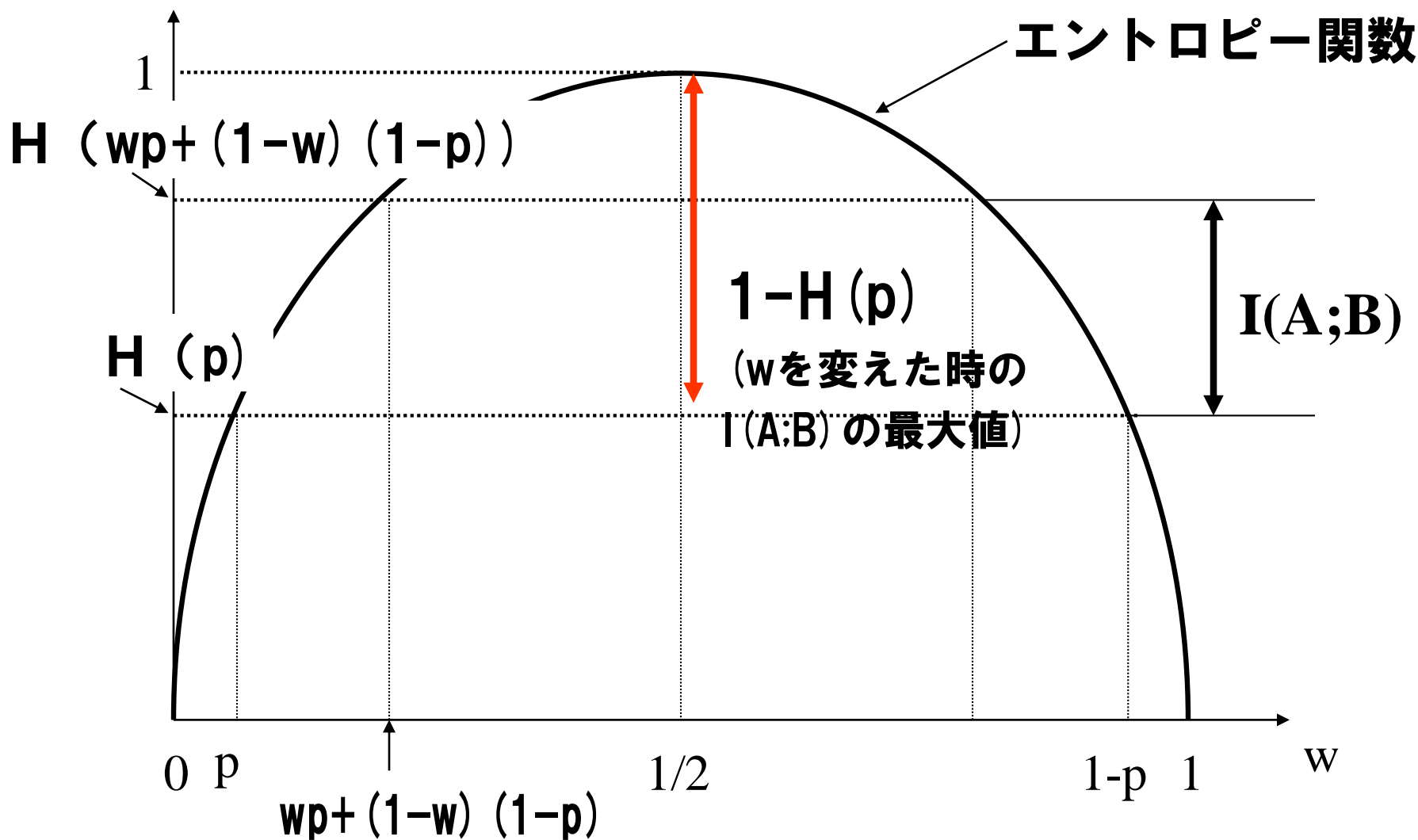

第1項


第2項

- 第1項 $H(B)$ は Bの分布のみに依存する。
- 第2項は 通信路行列のみに依存する。

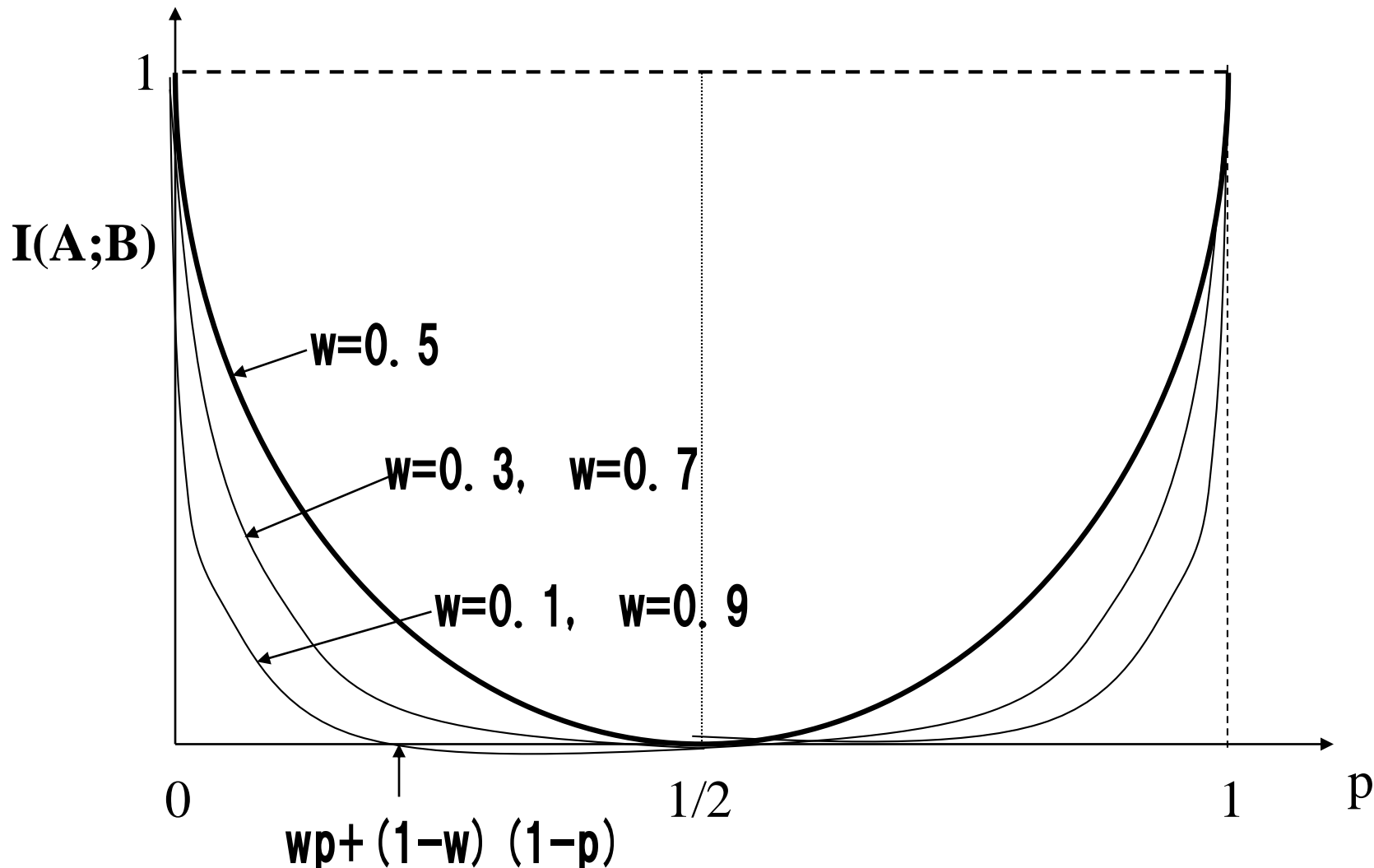
8.9 BSCの相互情報量：グラフ表現

入力シンボルの発生確率 w の関数としての図



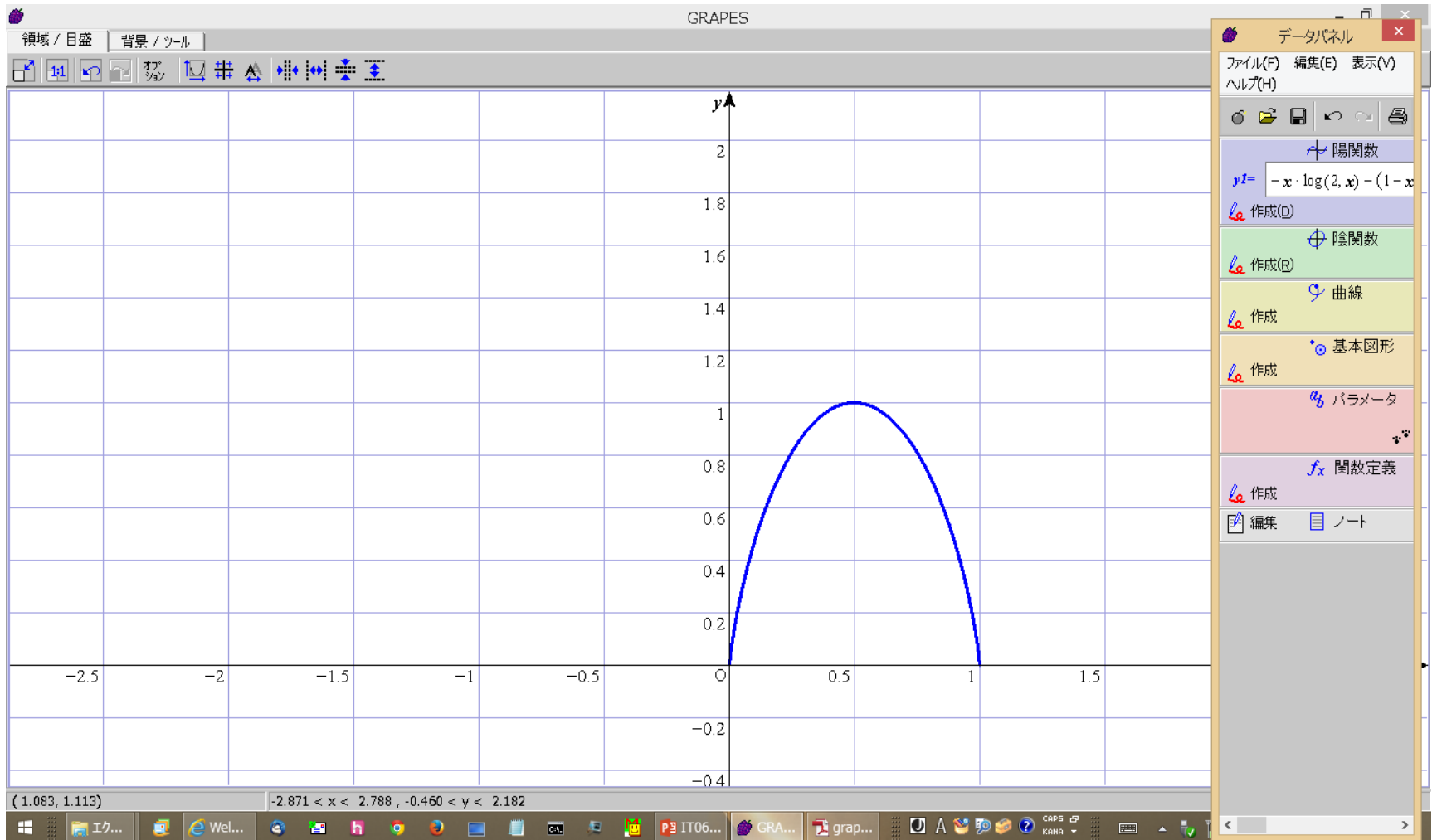
8.9 BSCの相互情報量：グラフ表現

通信路のシンボル誤り率 p の関数としての図



関数グラフ作成ソフト GRAPES

<http://www.osaka-kyoiku.ac.jp/~tomodak/grapes/>



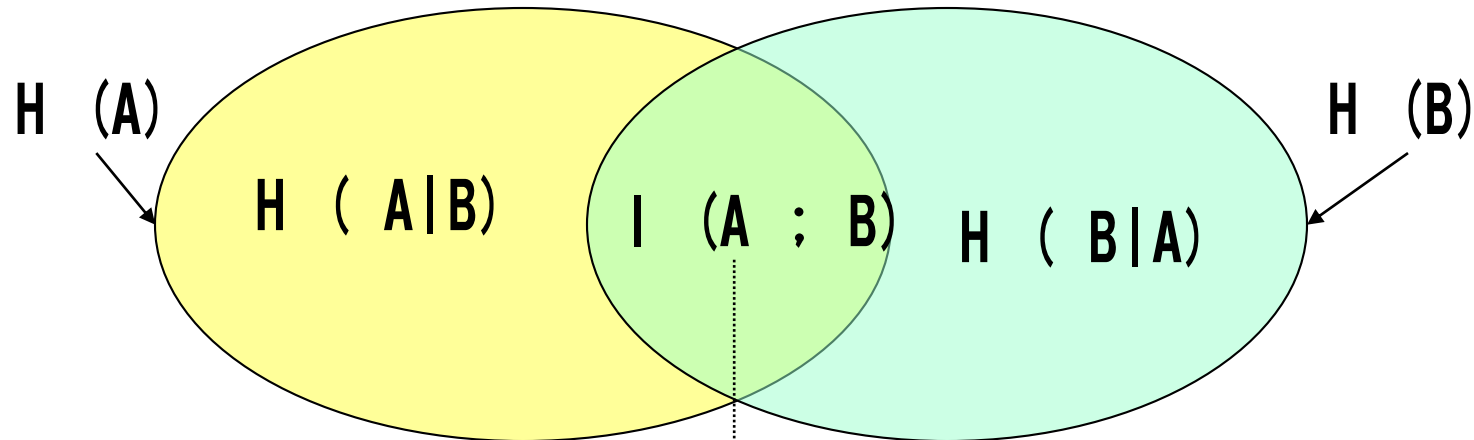
8.10 相互情報量の凸性

- 相互情報量はBSCに限らず凸性をもつ
- $I(X; Y) = H(Y) - H(Y|X)$ の関係があり、 $H(Y)$ は入力 (X) の確率変数 w に関して上に凸であり、第2項 $H(Y|X)$ も w に関して線形であることから、全体として上に凸である。
- また、通信路の確率変数 p に関しては、下に凸であることが証明される（省略）。

8.11 相互情報量 $I(A; B)$ の性質

- $I(A; B) = \sum_{A, B} P(a_i, b_j) \log [P(a_i, b_j) / P(a_i) P(b_j)]$ より、 $I(A; B)$ は A と B に関して対称的
 - $I(A; B) = I(B; A) = H(B) - H(B|A)$
 - $I(A; B) = H(A) - H(A|B) = H(B) - H(B|A)$
 - $I(A; B) = H(A) + H(B) - H(A, B)$
但し、 $H(A, B) = -\sum_{A, B} P(a_i, b_j) \log P(a_i, b_j)$ 、は結合エントロピー
- $0 \leq I(A; B) \leq H(A)$: 非負で $H(A)$ 以下
- $0 \leq I(A; B) \leq H(B)$: 非負で $H(B)$ 以下
- A と B が独立の事象の時、 $I(A; B) = 0$
- 雑音のない通信路の時、 $I(A; B) = H(A)$

8.11 相互情報量 $I(A; B)$ の性質



$$I(A; B) = H(A) + H(B) - H(A, B)$$

セミコロンに注意！
(相互であることを示す)

コンマに注意！
(結合事象のエントロピーであることを示す)

- $H(A)$, $H(B)$: 事象系A, Bに関するエントロピー
- $H(A|B)$: Bを知った時のAに関するエントロピー
- $H(B|A)$: Aを知った時のBに関するエントロピー
- $H(A, B)$: 結合事象系 ($A \times B$) に関するエントロピー

8.11 相互情報量 $I(A; B)$ の補足

$I(A; B)$ は:

- Bを知ることにより、Aを表現するのに必要なビット数が何ビット節約できるかを示している。
 - $H(A) - H(A|B)$
- AとBに共通に含まれる情報量
 - Bを知ることにより、Aに関してはBの中に含まれるAの情報量だけ「知識」が増える。この増加分だけ、Bを知った後ではAに関するあいまい性が減り、必要なビット数が節約できる。
- したがって、AとBに関して対称的
 - $H(B) - H(B|A) = I(B; A)$ でもある

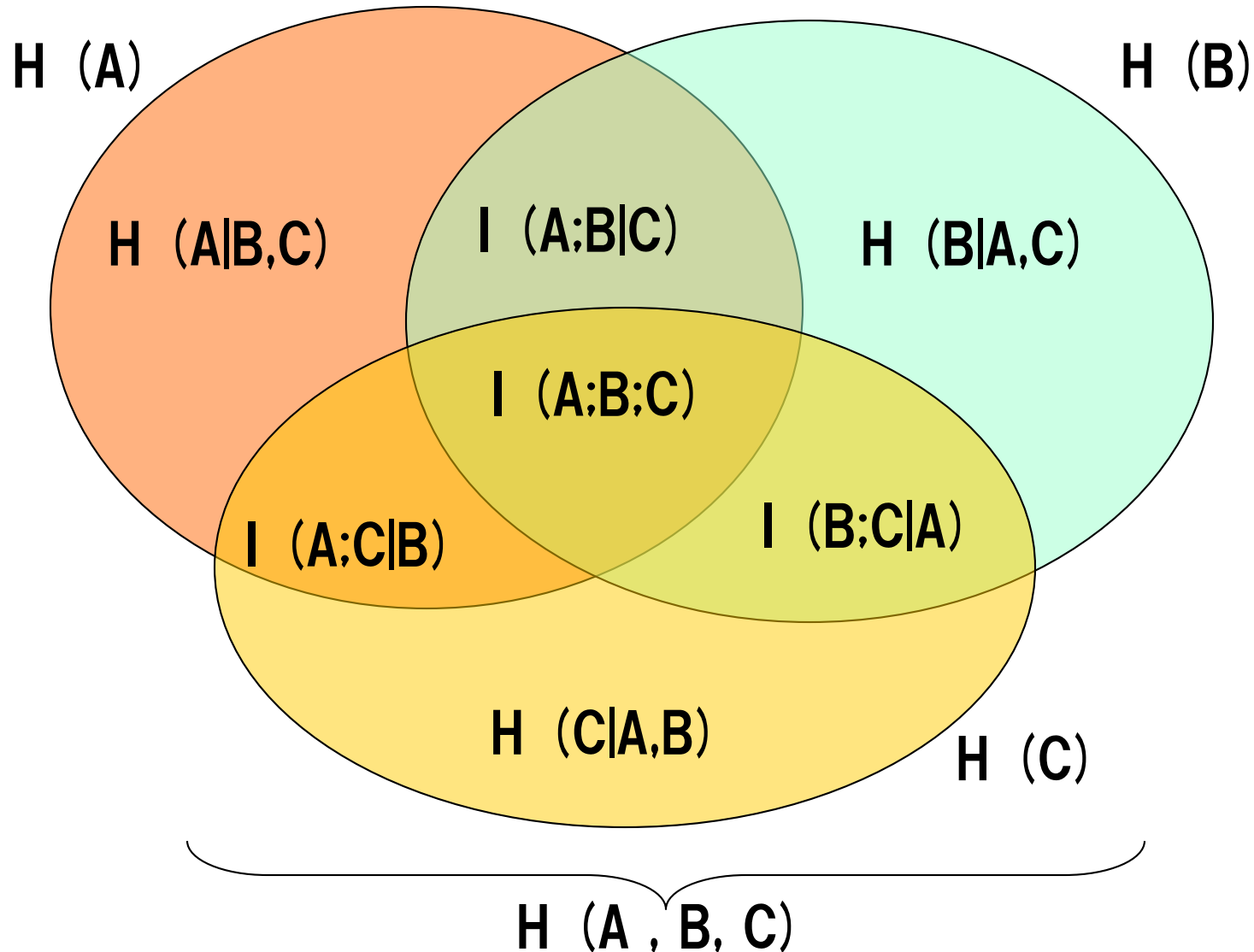
8.11 相互情報量 $I(A; B)$ の性質:加法性

- 3つ以上の事象系に拡張することができる
- 条件付相互情報量 $I(A; B|C)$
 - **通報Cを知った後での、AとBの相互情報量**
 - Aに関する2通りの通報 B, Cがあり、既に通報Cの内容が分かっている時、新たに通報 Bより得られる情報量:
 $H(A|C) - H(A|B, C)$
 - AとBが共通にもつ情報量から、A, B, Cが共通にもつ情報量を引いたもの: $I(A; B) - I(A; B; C)$
- 条件付相互情報量 $I(A; C|B)$
 - **通報Bを知った後での、AとCの相互情報量**
 - Aに関する2通りの通報 B, Cがあり、既に通報Bの内容が分かっている時、新たに通報 Cより得られる情報量:
 $H(A|B) - H(A|B, C)$

8.11 相互情報量 $I(A; B)$ の性質:加法性

- AとCが共通にもつ情報量から、A,B,Cが共通にもつ情報量を引いたもの: $I(A;C) - I(A;B;C)$
- 条件付相互情報量 $I(B;C|A)$
 - **通報Aを知った後での、BとCの相互情報量**
 - Bに関する2通りの通報 A, Cがあり、既に通報Aの内容が分かっている時、新たに通報 Cより得られる情報量:
 $H(B|A) - H(B|A, C)$
 - BとCが共通にもつ情報量から、A,B,Cが共通にもつ情報量を引いたもの: $I(B;C) - I(A;B;C)$
- A,B,Cの共通の相互情報量 $I(A;B;C)$
 - $I(A;B;C) = I(A;B) - I(A;B|C)$
 $= H(A) + H(B) + H(C) - H(A, B) - H(A, C)$
 $\quad - H(B, C) + H(A, B, C)$
I(A;B)等は常に非負であるが、 $I(A;B;C)$ は負になることもある

8.11 相互情報量 $I(A; B)$ の性質: 加法性



8.12 3事象間の相互情報量 $I(A;B;C)$ 例1

- $A=B=C = \{0, 1\}$ として、結合確率 $P(a,b,c)$ が
 - $P(a=0,b=0,c=0) = 1/4,$
 - $P(a=1,b=0,c=1) = 1/4,$
 - $P(a=0,b=1,c=1) = 1/4,$
 - $P(a=1,b=1,c=0) = 1/4$
 - その他の $P(a,b,c) = 0$ とすると、
 - $I(A;B;C) = I(A;B) - I(A;B|C)$
 $= H(A) + H(B) + H(C) - H(A,B) - H(A,C)$
 $- H(B,C) + H(A,B,C)$ より求めると、
 - $I(A;B) = 0, I(A;B|C) = 1, I(A;B;C) = -1$

8.12 $I(A;B;C)$ 例1の解釈

- AとBが全くランダムに0、1をとり、CはAとBの2を法とする和で与えられる場合に相当する。
- AとBはランダムなので、Bを知ってもAに関する情報は得られない。従って、 $I(A;B)=0$ 。
- しかし、例えば $C=1$ ということを知っていると、Bから2を法とする和を計算することにより、確実にAが推定できる。

8.12 3事象間の相互情報量:例2

YZ	晴寒 (y1z1)	晴暖 (y1z2)	雨寒 (y2z1)	雨暖 (y2z2)
X				
繁 (X1)	0.12	0.08	0.27	0.16
閑 (X2)	0.06	0.15	0.04	0.12

- ある都市の天候 Y (晴/雨)、温度 Z (寒/暖) と交通量 X (閑/繁) の関係を調査した結果上記の分布をした (数字は相対頻度を確率で表したもの)
- 天候 Y と温度 Z が知らされたとき、交通量 X に関してどれだけ情報が得られるか?
- 求める値は平均相互情報量 $I(X; YZ)$
 - $I(X; YZ) = I(X; Y) + I(X; Z|Y)$ の関係が成立する

8.12 3事象間の相互情報量:例2の計算

- $I(X;Y) = p(x_1, y_1) \log \{p(x_1, y_1) / p(x_1) p(y_1)\} + \dots + p(x_2, y_2) \log \{p(x_2, y_2) / p(x_2) p(y_2)\}$
 $= 0.2 \log(0.2 / (0.63 * 0.41)) + 0.43 \log(0.43 / (0.59 * 0.63)) + 0.21 \log(0.21 / (0.41 * 0.37)) + 0.16 \log(0.16 / (0.59 * 0.37)) = 0.04$
- $I(X;Z|Y) = p(x_1, y_1, z_1) \log \{p(z_1|x_1, y_1) / p(z_1|y_1)\} + \dots + p(x_2, y_2, z_2) \log \{p(z_2|x_2, y_2) / p(z_2|y_2)\}$
 $= 0.12 \log \{0.6 / (0.18 / 0.41)\} + 0.08 \log \{0.4 / (0.23 / 0.41)\} + 0.27 \log \{(0.27 / 0.43) / (0.31 / 0.59)\} + 0.16 \log \{(0.16 / 0.43) / (0.28 / 0.59)\} + 0.06 \log \{(0.02 / 0.07) / (0.18 / 0.41)\} + 0.15 \log \{(0.05 / 0.07) / (0.23 / 0.41)\} + 0.04 \log \{(1/4) / (0.31 / 0.59)\} + 0.12 \log \{0.75 / (0.28 / 0.59)\} = 0.08$
- $I(X;YZ) = 0.04 + 0.08 = 0.12$
- $H(X) = -p(x_1) \log p(x_1) - p(x_2) \log p(x_2)$
 $= -0.63 \log 0.63 - 0.37 \log 0.37 = 0.95$
- **従って、元々交通状態がもつ情報量0.95に対して、天候、寒暖情報を知ったときに得られる情報量0.12は、わずか13%しかなく、より多くの情報量を得るためには、その他の事象を観測することが必要である、ことを意味している。**