

情報理論 (No. 2)

2016/9/24

今後の講義予定

10／1 : 第3回

10／8 : 第4回

10／15 : 第5回

10／22 : 平塚祭のため休み

講義目次

- **2. 情報量**
 - 2.1 情報量の定義
 - 2.2 情報量の計算
- **3. 情報源**
 - 3.1 無記憶情報源
 - 3.2 平均情報量（エントロピー）
 - 3.3 無記憶情報源の拡大
 - 3.4 マルコフ情報源

2. 情報量



2.1 情報量の定義

2.2 情報量の計算

2.1 情報量の定義：組合せ論的情報量

- 組合せ論的情報量

- 情報は「未知のものを知ること得られるもの」
- 最も単純な情報は、「あり／無しの可能性から1つを確定する情報」。これより、情報の基本単位を次のように定義する。
- 情報の基本単位(1情報量)=2つの場合から1つを確定する情報
- この考えを進めて、情報量=多くの場合の中から1つを確定する情報=予想される答の数、とする。
 - この定義によれば、Yes/Noで答えられる問の数が n 個の場合、予想される答の数 $k=2^n$ となる。従って、次の関係が成り立つ：
 - $n=\log_2 k$ （ k 個の場合から1つを確定する情報）
 - $k=2$ ならば $n=1$ 、 $k=4$ ならば $n=2$ 、 $k=8$ ならば $n=3$
- このような情報量の定義を組合せ論的情報量という。

2.1 情報量の定義：確率的情報量

- 確率を基に情報量を定義する
 - 組合せ数が2のべき乗の事象以外等への自然な拡張により、確率論をベースにした情報量を定義する
- 確率的な情報量の定義
 - 『生起確率 $p(a)$ の事象が起こったとき、これを知ることのでられる情報量 $I(a)$ を、確率の逆数の対数、すなわち、
$$I(a) = \log(1/p(a)) = -\log p(a)$$
で定義する』 これを事象 a の自己情報量という。
 - 対数の底により、単位が変わる
底=2 (ビット bit)、10 (ディット dit)、自然対数 e (ナット nat)
 - $p=1/2=0.5$ の時、 $I(a)=1$ (ビット)。これを基本情報量という。

2.1 情報量の定義：確率的情報量

- 上の定義理由の説明

$$I = f(p)$$

情報量は次の性質を満たす必要がある。

(性質1) **逆単調性**： $p_1 < p_2$ ならば $f(p_1) > f(p_2)$

珍しいことのほうが情報量が多い

(性質2) **加法性**： 事象 $E_1 (p_1)$ 、 $E_2 (p_2)$ が独立ならば、
 $f(p_1 \cdot p_2) = f(p_1) + f(p_2)$

確率が「積」のときは情報量は「和」になる

例：事象 E_1 『9月10日東京は晴だった』

事象 E_2 『9月10日のA社の株価が1000円だった』

は独立事象。両者を知った情報量は各々の情報量の和になる。

(性質1) (性質2) を満たす関数は、 $f(p) = -\log p$ となる。

⇒次の定理を参照

2.1 情報量の定義：確率的情報量

【定理】 I を p に関する関数 $I(p)$ とするとき、

(1) $I(p)$ は p の単調減少の連続関数 $\dots (2.1-1)$

(2) $p=p_1 * p_2$ とするとき、 $I(p)=I(p_1 p_2)=I(p_1)+I(p_2)$ $\dots (2.1-2)$

の両方の条件を満たす関数は、 $I(p)=-k \log_2 p$ に限られる。
ここで、 k は正の定数である。

【証明】 M を正の整数として、

$$p=p_1^M \quad \dots (2.1-3)$$

とする。ここで上記の条件(2.1-2)を用いると、

$$I(p)=I(p_1 p_1^{M-1})=I(p_1)+I(p_1^{M-1}) \quad \dots (2.1-4)$$

となる。 $I(p_1^{M-1})$ も同様に展開すると、結局

$$I(p_1^M)=M I(p_1) \quad \dots (2.1-5)$$

となる。

式(2.1-2)に、 $p_1=p_2=p=1$ を適用すると、 $I(1)+I(1)=I(1)$

が得られるので、 $I(1)=0$ $\dots (2.1-6)$

2.1 情報量の定義：確率的信息量

また、 $I(p)$ は単調減少関数なので、

$$I(p) > 0 \quad (p > 0) \quad \dots (2.1-7)$$

も導かれる。

更に M, r を2より大きな正の整数とすると、 2^r を M^N と M^{N+1} の間に含むような正の整数 N が必ず存在する。



$$\text{即ち、} 1/M^N \geq 1/2^r > 1/M^{N+1} \quad \dots (2.1-8)$$

これらを引数にもつ単調減少関数 $I(p)$ は、

$$I(1/M^N) \leq I(1/2^r) < I(1/M^{N+1}) \quad \dots (2.1-9)$$

なる関係が得られる。これに式(2.1-5)の関係を適用すると

$$N \cdot I(1/M) \leq r \cdot I(1/2) < (N+1) \cdot I(1/M) \quad \dots (2.1-10)$$

が得られる。この式を変形すると、

2.1 情報量の定義：確率的信息量

$$\frac{N}{r} \leq \frac{I(\frac{1}{2})}{I(\frac{1}{M})} < \frac{N+1}{r} \quad (2.1-11)$$

一方、式(2.1-8)の対数をとって式を変形すると、

$$\log \frac{1}{M^N} \geq \log \frac{1}{2^r} > \log \frac{1}{M^{N+1}} \quad (2.1-12)$$

$$M \log \frac{1}{M} \geq -r > (N+1) \log \frac{1}{M} \quad (2.1-13)$$

$$\frac{N}{r} \leq \frac{1}{-\log \frac{1}{M}} < \frac{N+1}{r} \quad (2.1-14)$$

式(2.1-11)と式(2.1-14)から、

$$\frac{1}{-\log \frac{1}{M}} \text{ と } \frac{I(\frac{1}{2})}{I(\frac{1}{M})} \text{ は、共に区間 } \left(\frac{N}{r}, \frac{N+1}{r} \right) \text{ の内にあることがわかる。}$$

2.1 情報量の定義：確率的信息量

従って、次の式が成り立つ。

$$\left| \frac{1}{-\log \frac{1}{M}} - \frac{I(\frac{1}{2})}{I(\frac{1}{M})} \right| < \frac{1}{r} \quad (2.1-15)$$

ここで、**M**は固定し、**r**を任意に大きくできるので、 $r \rightarrow \infty$ とすると、左辺は**0**になるので、

$$I\left(\frac{1}{M}\right) = -I\left(\frac{1}{2}\right) \log \frac{1}{M} \quad (2.1-16)$$

これは、変数**P**がある正の整数**M**の逆数で表される($P=1/M$)ときは、

$$I(P) = -I\left(\frac{1}{2}\right) \log P \quad (2.1-17)$$

と表されることが分かる。**P**が有理数のときは、適当な整数**M**, **N**を用いて、 $P=N/M$ と表される。

2.1 情報量の定義：確率的情報量

$$\frac{1}{M} = \frac{N}{M} \cdot \frac{1}{N} \quad (2.1-18)$$

式(2.1-18)に条件式 (2.1-2)を適用すると、次の式が成り立つ。

$$I\left(\frac{1}{M}\right) = I\left(\frac{N}{M}\right) + I\left(\frac{1}{N}\right) \quad (2.1-19)$$

ここで、 $I(1/M), I(1/N)$ に対して式(2.1-16)が適用できるので、

$$-I\left(\frac{1}{2}\right)\log \frac{1}{M} = I\left(\frac{N}{M}\right) - I\left(\frac{1}{2}\right)\log \frac{1}{N} \quad (2.1-20)$$

と変形でき、次のようになる。

$$I\left(\frac{N}{M}\right) = -I\left(\frac{1}{2}\right)\log \frac{N}{M} \quad (2.1-21)$$

このように、 P が有理数のときも式(2.1-17)が成り立つ事が分かる。

2.1 情報量の定義：確率的信息量

- P が無理数のときも、有理数にいくらでも近似できるので、 P の連続性より、**任意の P に対して、定理が成り立つ事が証明できる。**

2.2 情報量の計算

以後、特に断らないかぎり単位はビット (bit) とする

- **トランプカード 52 枚に関する情報量**

- ハートの A であることを知った時の情報量

$$I(\text{ハート A}) = -\log_2(1/52) \doteq 5.70$$

- ハートであることのみ知った時の情報量

$$I(\text{ハート}) = -\log_2(1/4) = 2$$

- A (エース) であることのみ (種別関係なく番号を) 知った時の情報量

$$I(A) = -\log_2(1/13) \doteq 3.70$$

- **性質：加法性**

$$I(\text{ハート A}) = I(\text{ハート}) + I(A)$$

カードの種別とカードの番号は互いに独立(依存関係がない)ので、
この加法性が成り立つ

2.2 情報量の計算（その1）

問題1. チームAとBの決勝戦（引分け無し）で、Aが勝つ確率=0.4、Bが勝つ確率=0.6のとき、それぞれが勝った時に、それを知ることによって得られる情報量($I(A)$, $I(B)$)を計算せよ。

（答） $I(\text{勝}=A) = -\log 0.4 \doteq 1.3$; $I(\text{勝}=B) = -\log 0.6 \doteq 0.7$

問題2. ある人が長い睡眠から覚め、ラジオの時報（毎時間に1回）の音を聞いた。この人が今の時刻を教えてもらうことによって得られる情報量はいくらか。

（答）時刻を教えられない場合は、時報は24回ある（ラジオで毎時放送すると仮定する）。教えられた場合は、そのうちのどれかが確定するので、 $I(\text{時報}) = -\log (1/24) \doteq 4.6$

2.2 情報量の計算（その2）

問題3.（同じ形のものを）天秤で計ったとき得られる情報量はいくらか。

（答）天秤で重さを比較する試行は、「右が重い」「左が重い」「両者同じ」、の3通りの結果しか持たない。従って、1回の試行によって得られる情報量 $I(\text{天秤}) = -\log_2(1/3) \doteq 1.6$ 以下となる。

問題4. 同じ形の7個の玉があり1つだけ軽い玉が混じっている。これを何回の天秤計量すれば確実に見つけれられるか。

（答）2回。7個を（3個、3個、1個）にわけ、3個同士を天秤にかける。もし釣合えば、残り1個が軽いとわかる。釣合わなければ、軽い方の3個のうち、2個を選び天秤にかける。釣合えば残り1個が軽いし、釣合わなければ軽いほうがわかる。結局、2回で可能。

情報量的には、次のように計算できる： 1回の試行で得られる情報量は $\log_2 3$ で、 k 回の試行では $k \log_2 3$ となる。7個の玉のうち1個が軽い確率は $1/7$ なので情報量は $\log_2 7$ となり、 $k \log_2 3 \geq \log_2 7$ となるような最小の整数 k は2となる。

3. 情報源

3.1 無記憶情報源

3.2 平均情報量（エントロピー）

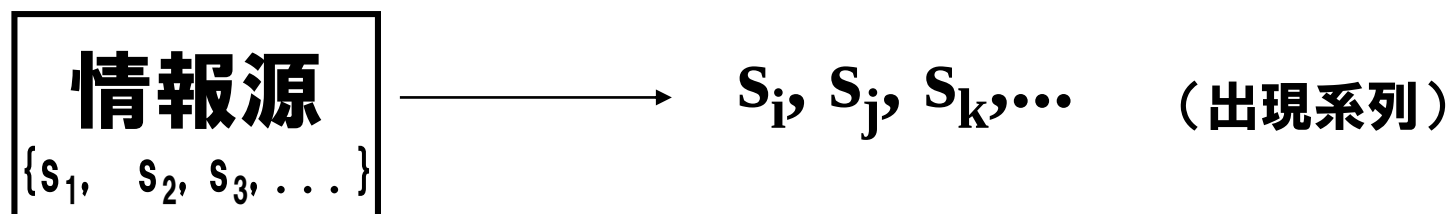
3.3 無記憶情報源の拡大

3.4 マルコフ情報源

これから、本格的に数学を使う

3.1 無記憶情報源：最も基本的な情報源

個々の事象でなく、事象が次々と発生する発生源（情報源という）を考えて、情報源の性質や量を定義する



情報源アルファベット $S = \{s_1, s_2, s_3, \dots, s_q\}$

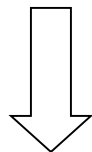
(例)情報源＝トランプカード

- アルファベット $\{s_1=\text{スペードA}, s_2=\text{スペード2}, \dots, s_{14}=\text{ハートA}, \dots, s_{52}=\text{クラブK}\}$
($s_1 \sim s_{13}$: スペード、 $s_{14} \sim s_{26}$: ハート、 $s_{27} \sim s_{39}$: ダイヤ、 $s_{40} \sim s_{52}$: クラブ、とする)
- 出現系列の例=(スペード8, ハート3, ダイヤA,...)= $s_8, s_{16}, s_{27}, \dots$

各出現シンボル s_i, s_j が 統計的に独立 であるとき、情報源は無記憶 であるという。

3.1.1 無記憶情報源の表現と情報量

$$S = \left[\begin{array}{cccc} s_1, & s_2, & s_3, & \dots, & s_n \\ P(s_1), & P(s_2), & P(s_3), & \dots, & P(s_n) \end{array} \right]$$



事象 s_3 が起きた場合に受取る情報量 $I(s_3)$ は、

$$I(s_3) = \log_2 1/P(s_3) \quad (\text{ビット})$$

これを、全ての事象について平均をとると
1シンボル当たりの情報量が出る

3.2 平均情報量（エントロピー）

3.2.1 エントロピーの定義

- **平均情報量**は次の式で定義される
 - 平均情報量を**エントロピー**とも呼ぶ。

$$H(S) = \sum P(s_i) \log_2 1/P(s_i)$$

（単位=ビット／シンボル）

3.2.2 エントロピーの計算式

$$H(S) = \sum P(s_i) \log_2 1/P(s_i)$$

添字について総和をとる（単位=ビット／シンボル）

上式は、下記のように展開できる。

- $H(S) = -P(s_1) \log P(s_1) - P(s_2) \log P(s_2) - \dots - P(s_n) \log P(s_n)$

- これは、各事象のもつ情報量, $-\log P(s_i)$, にその発生する確率 $P(s_i)$ を掛けて平均値を求めている。

一般的に、確率的な事象を要素とする集合の平均値(期待値)は、各確率変数 (s_i) がとる値 $(-\log P(s_i))$ に発生確率 $p(s_i)$ をかけて得られる。

3.2.3 エントロピーの計算例

- 9月10日の横浜の天気は、雨、晴、曇、雪の4種類の可能性があり、次の確率で与えられる場合：

$$S(\text{天気}) = \{s_1=\text{雨}, s_2=\text{晴}, s_3=\text{曇}, s_4=\text{雪}\}$$

$$P(s_1) = 1/2, P(s_2) = 1/4, P(s_3) = 1/4, P(s_4) = 0$$

天気を知らせる通報より得られる情報量は、

$$H(S) = -1/2 \log 1/2 - 1/4 \log 1/4 - 1/4 \log 1/4 - 0 * \log 0 = 1.5$$

(ビット/シンボル)

- これは横浜の天気をもつ平均情報量とも考えることができる
- $0 * \log 0 = 0$ になることに注意

3.2.4 エントロピーの意味・解釈

■ 意味

- $H(S)$ は情報源の**1シンボルが平均としてもつ情報量**
- 情報源の曖昧性 $H(S)$ が通報を受取ることによってゼロになる
 - ある s_i の $p(s_i) = 1$, その他 $p(s_j) = 0$ になるため

■ 解釈

- 情報源シンボルの**不確実度、曖昧さ、無秩序さ**の尺度
- 元は統計力学における気体分子のエントロピーが、

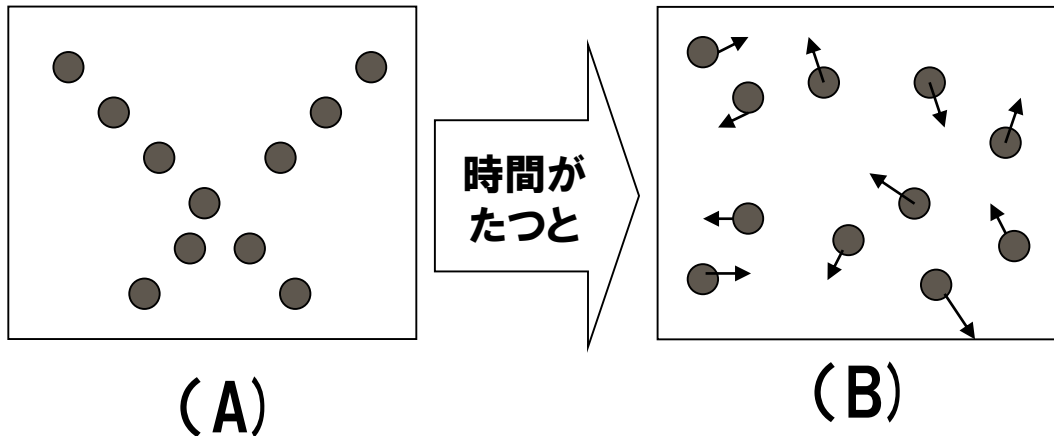
$$H = -B \sum_k N_k \log_e N_k \quad (k \text{ について総和をとる}),$$
$$B = \text{Boltzman定数}, N_k = \text{気体分子の} k \text{ 番目エネルギー状態にある確率}$$

で与えられることに由来する。

情報理論のエントロピーは統計力学と密接な関係がある

3.2.5 物理学のエントロピーの直感的意味

- **気体や液体の分子の運動状態の無秩序さ**
 - 無秩序であるほうがエントロピーは大きい
- **熱力学の基本法則の1つ（熱力学第2法則）**
「**外界と閉じた系の中でエントロピーは増大する**」
（即ち下図の（B）のような系に次第に移る）

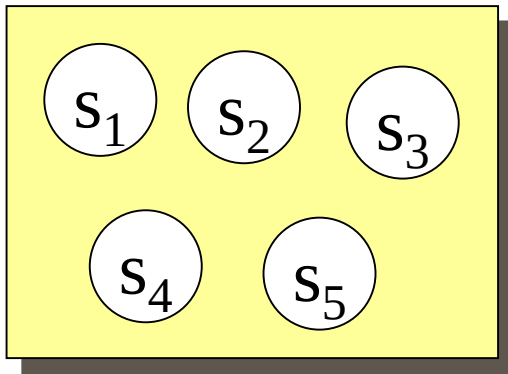


(B)の方が(A)より無秩序
性が高いのでエントロ
ピーも大きい

Aのエントロピー < Bのエントロピー

3.2.6 平均情報量とエントロピーの関係

情報源 (S)



例：S = ある場所の天気種別

$S_1 = \text{晴}$ 、 $S_2 = \text{曇}$ 、 $S_3 = \text{雨}$ 、 $S_4 = \text{霧}$ 、 $S_5 = \text{雪}$

通報Mとは例えばある期間の天気事象を表す

$M = \{1 \text{ 日は } s_1, 2 \text{ 日は } s_3, 3 \text{ 日は } s_4, \dots\}$

平均情報量

エントロピー

通報Mから平均的に得られるニュース価値

=

通報Mを知らない状態で情報源が持つ不確かさ

情報は負の不確かさ(エントロピー)と考えられるため、
ネグエントロピー (negentropy=negative entropyに由来) と呼ぶことがある。これは熱力学における物質の乱雑さの度合いを示す物理量でもある。

3.2.7 エントロピーの計算問題（例）

- **問題**：ある集団の100人（＝事象系Sとする）に、購読新聞を聞いたら、以下ようになった。新聞は5種類しかないと仮定し、この比率は全集団で同じとすると、この集団（S）の購読新聞に関するエントロピーはいくらか。（ $\log_2 10 = 3.32$ ）
25人――読売
10人――毎日
50人――朝日
5人――日経
10人――神奈川
- **解答**：E1=ある人が読売購読者、E2=同様に毎日購読者、E3=同様に朝日購読者、E4=同様に日経購読者、E5=同様に神奈川購読者、とする。この情報源(S)は、 $p(E1)=0.25$, $p(E2)=0.1$, $p(E3)=0.5$, $p(E4)=0.05$, $p(E5)=0.1$ であるので、情報源がもつ平均自己情報量（エントロピー）は、定義により、
$$H(S) = p(E1) \log (1/p(E1)) + p(E2) \log (1/p(E2)) + p(E3) \log (1/p(E3)) + p(E4) \log (1/p(E4)) + p(E5) \log (1/p(E5))$$
$$= 0.25 \log 4 + 0.1 \log 10 + 0.5 \log 2 + 0.05 \log 20 + 0.1 \log 10 = \underline{1.88} \quad (\text{単位はビット／シンボル})$$

3.2.8 エントロピーの上限・下限

- 平均情報量の上限、下限に関する定理：

$H(S)$ の上下限は、 S に含まれる事象（記号）数が n のとき、 $0 \leq H(S) \leq \log_2 n$ で与えられる。

下限は、ある s_i の $p(s_i) = 1$, $p(\text{他の } s_j) = 0$ のとき成立。

上限は、全ての事象が等しい確率のとき、即ち $p(s_i) = 1/n$ のとき成立。

- 証明：
 - 対数関数の不等式と補助定理を使い証明できる
 - 詳細は次ページ以下

3.2.9 エントロピーの上限・下限の証明： 対数関数の不等式と補助定理

- 対数関数の不等式

$\ln x \leq x - 1, x > 0$ 、等号は $x = 1$ のとき成立

$\ln x$ は、 x の自然対数(e を底)を表す関数($\log_e x$)を表わす。

$\ln x = \log x / \log e$ の関係がある。

この不等式は、グラフを描くことで示される(次ページ)。

この不等式を使って次の補助定理が証明できる。

- 補助定理(シャノンの補助定理)

$p_1, p_2, \dots, p_n$ ($p_i \geq 0, \sum p_i = 1$), $q_1, q_2, \dots, q_n$ ($q_i \geq 0, \sum q_i = 1$) を2通りの確率分布とする。このとき、以下の不等式が成立する：

$$-\sum p_i \log p_i \leq -\sum p_i \log q_i$$

(等号は $p_i = q_i$ のとき)

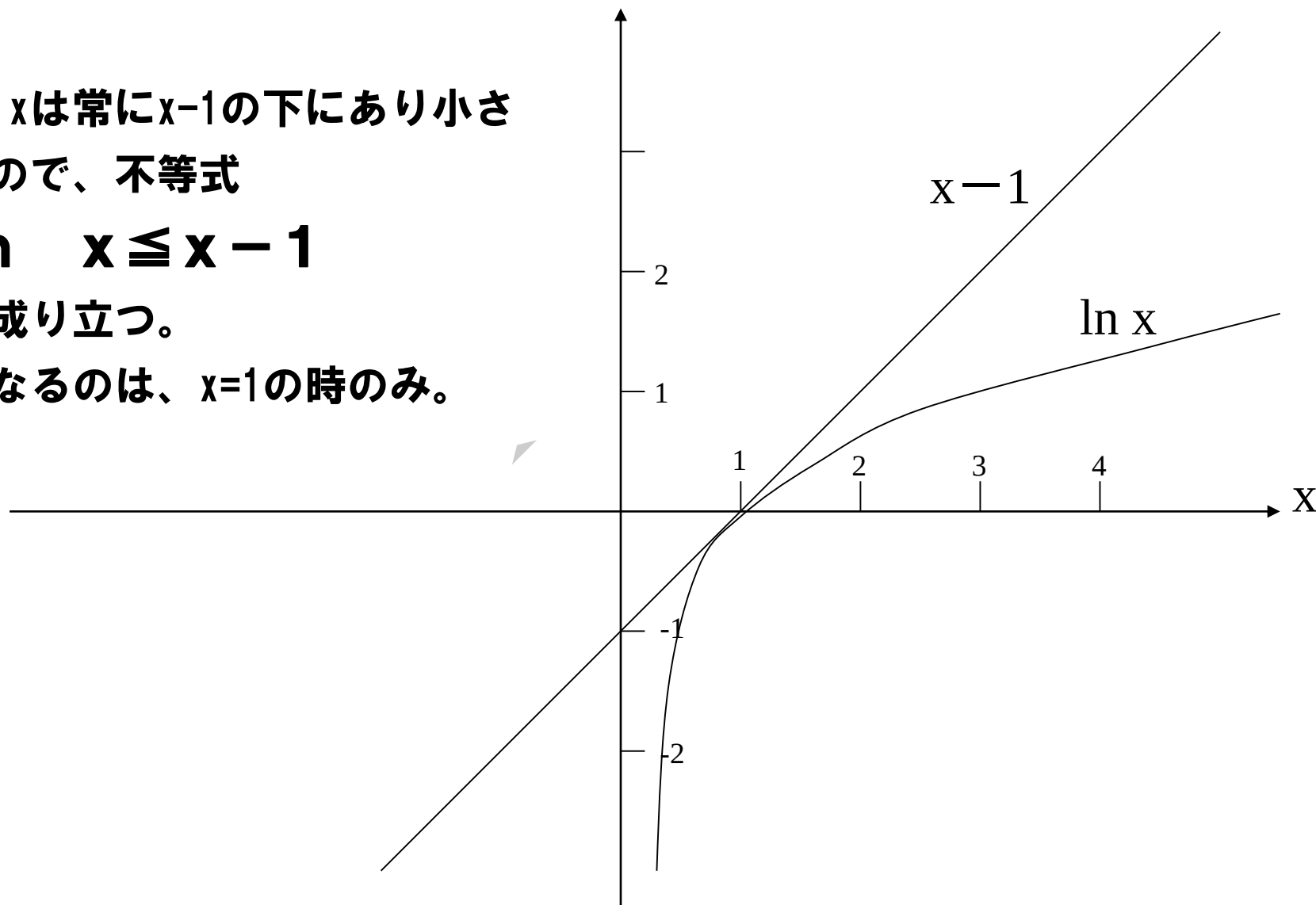
3.2.9 エントロピーの上限・下限の証明： 関数 $\ln x$ と $x-1$ の関係

$\ln x$ は常に $x-1$ の下にあり小さいので、不等式

$$\ln x \leq x - 1$$

が成り立つ。

重なるのは、 $x=1$ の時のみ。



3.2.9 エントロピーの上限・下限の証明： 補助定理の証明

- $-\sum p_i \log_2 p_i + \sum p_i \log_2 q_i = (1/\ln 2) \sum p_i \ln (q_i/p_i)$ が成り立つ。

(証明) 左辺の式 $= -\sum p_i (\log_2 p_i - \log_2 q_i) = \sum p_i (\ln q_i / \ln 2 - \ln p_i / \ln 2) =$ 右辺の式

- この右辺に不等式、 $\ln x \leq x-1$ を適用する

$$\begin{aligned} (1/\ln 2) \sum p_i \ln (q_i/p_i) &\leq (1/\ln 2) \sum p_i (q_i/p_i - 1) \\ &= (1/\ln 2) (\sum q_i - \sum p_i) = 0 \end{aligned}$$

- 結局、 $-\sum p_i \log_2 p_i + \sum p_i \log_2 q_i \leq 0$ が成り立つので補助定理が証明された。
- また等号が成立するのは、 $q_i/p_i=1$, すなわち $q_i=p_i$ のときのみであることが分かる。

3.2.9 エントロピーの上限・下限の証明： 補助定理による証明

- **定理**

- $p_1, p_2, \dots, p_r$ を確率分布とする。即ち、 $p_i \geq 0, \sum p_i = 1$ 。このとき、
- $H(S) = -\sum p_i \log p_i \leq \log r$ が成立する。等号は、 $p_i = 1/r$ ($i=1, 2, \dots, r$) のときのみに限る。

- **証明**

- 補助定理において、 $q_i = 1/r$ となる確率分布を考えて代入する。
- $-\sum p_i \log p_i \leq \sum p_i \log r = \log r$

- **下限の証明**

- $H(S) \geq 0$ になることは、 $0 \leq p_i \leq 1$ から自明。

3.2.10 相対エントロピー、冗長度

- 出力シンボル数 r の情報源 S (r 元情報源) においては、エントロピーを $H(S)$ は最大値が $\log_2 r$ までいくので、一般に数が大きくなる。これを正規化した、

$$h(S) = H(S) / \log_2 r$$

を情報源 S の相対エントロピーという。

$0 \leq h(s) \leq 1$ を満たす。

- $r(S) = 1 - h(S)$

を情報源 S の冗長度という。

3.2.11 エントロピーの値の具体例： 英文を発生する情報源

- 英文を発生する情報源Sを考える（下図）
 - 英文を書くのに使う文字，記号は全部で27個（空白＋各アルファベットで，コンマやピリオドは除くとする）
 - これらの文字の生起確率は次ページ表のとおり
- エントロピーの計算

$$\begin{aligned} H(S) &= p(s_1) \log(1/p(s_1)) + p(s_2) \log(1/p(s_2)) + \cdots + p(s_{27}) \log(1/p(s_{27})) \\ &= 4.03 \text{ (bit/symbol)} \end{aligned}$$

情報源S（英語の文字，単語）

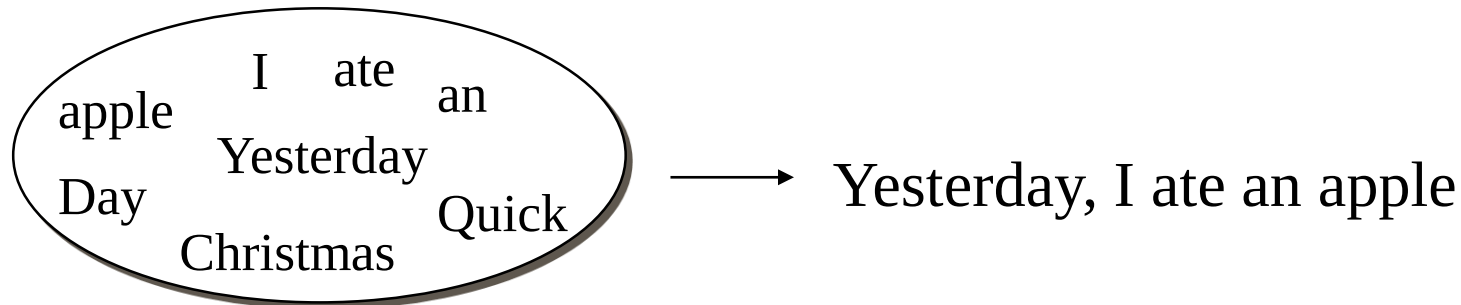


表 英語のアルファベットの生起確率

文字	生起確率	文字	生起確率	文字	生起確率
空白	0.186	I	0.057	R	0.048
A	0.064	J	0.001	S	0.051
B	0.013	K	0.005	T	0.080
C	0.022	L	0.032	U	0.023
D	0.032	M	0.020	V	0.008
E	0.103	N	0.057	W	0.017
F	0.021	O	0.063	X	0.001
G	0.015	P	0.015	Y	0.016
H	0.047	Q	0.001	Z	0.001

生起確率の大きい文字：

空白 > E > T > A > S > R > H > . . . > J, X, Z

例：英文を発生する情報源のエントロピー(2)

- エントロピーの計算（続き）

- もしこれらのシンボルが等確率で生じると仮定すれば、
1つのシンボルあたり、

$$\log 27 = 4.75 \text{ (bit/symbol)}$$

の情報量を有するはず。（即ちどのシンボルかを平等にしか当てられない場合は、 $1/27$ の確率になる）

- 相対エントロピーは、 $r=27$ に相当するので、

$$h(S) = H(S) / \log_2 r = H(S) / \log_2 27 = 4.03 / 4.75 = 0.848$$

- 冗長度は、

$$r(S) = 1 - h(S) = 1 - 0.848 = 0.152$$

となる。

3.2.12 無記憶 2 元情報源のエントロピー

- 無記憶2元情報源とは、2つの事象(s_1, s_2)だけが起こり、かつ独立に起きる情報源。各事象の生起確率が $p, 1-p$ で与えられる ($0 \leq p \leq 1$)

$$S = \begin{bmatrix} s_1 & s_2 \\ p & 1-p \end{bmatrix}$$

- この情報源のもつエントロピー：

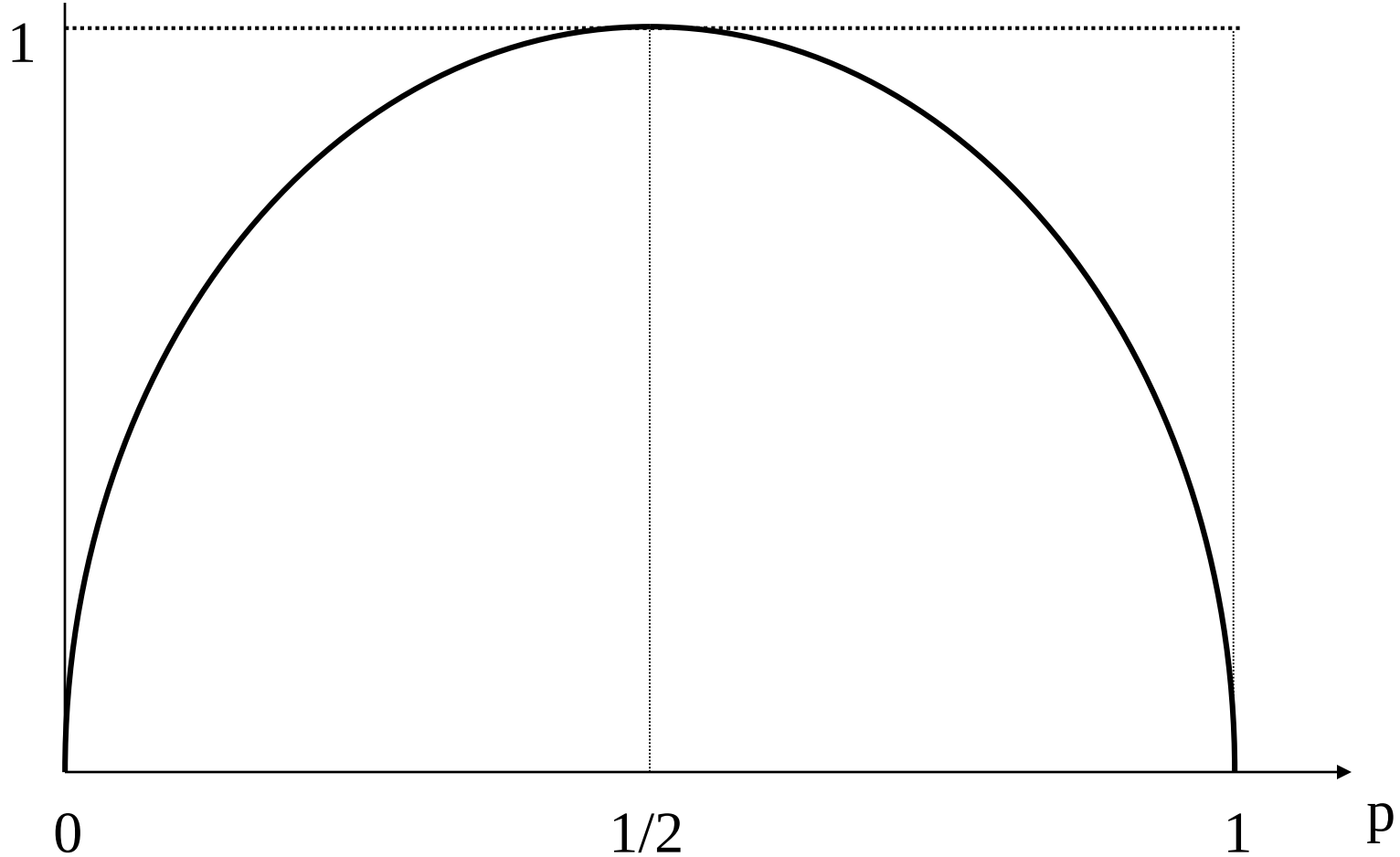
$$H(S) = -p \log p - (1-p) \log (1-p)$$

(p を変数とする関数と見たとき、**エントロピー関数**といい、 **$E(p)$** と表記される場合が多い)

単位は対数の底を2とすると「ビット／シンボル」

エントロピー関数 $E(p)$

$E(p)$



3.2.13 2元情報源のエントロピーの計算

- 問題

野球の試合が2つあり、それぞれ第1試合 (A), 第2試合 (B)とする。第1試合のチームは、巨人 (A_1)と阪神 (A_2)、第2試合のチームはヤクルト (B_1)と横浜 (B_2)とし、それぞれの勝つ確率が以下のように与えられるとする。

$$A = \begin{bmatrix} A_1 & A_2 \\ 7/8 & 1/8 \end{bmatrix} \quad B = \begin{bmatrix} B_1 & B_2 \\ 1/2 & 1/2 \end{bmatrix}$$

各試合のエントロピーを求め比較し、解釈せよ。

- 解答

$$\begin{aligned} H(A) &= -(7/8)\log(7/8) - (1/8)\log(1/8) = \\ &\quad -(7/8)\log 7 + (7/8)\log 8 + (1/8)\log 8 = 0.544 \end{aligned}$$

$$H(B) = -(1/2)\log(1/2) - (1/2)\log(1/2) = 1$$

従って、 $H(B) > H(A)$ 。確率分布が均質なものが曖昧性大きい。

3.2.14 結合エントロピー

- 2つの事象を同時に扱う事象系を**結合事象系**という
- 事象系Aの事象 a_i 、事象系Bの事象 b_j とする。この結合事象系の**結合確率**（同時に起きる確率）は $p(a_i, b_j)$ で表現される。
- 結合確率の性質
 1. $p(a_i, b_j) \geq 0$
 2. $\sum_i \sum_j p(a_i, b_j) = 1$ （すべての i, j について総和）
 3. $p(a_i) = \sum_j p(a_i, b_j)$ （添字 j のみについて総和）
 4. $p(b_j) = \sum_i p(a_i, b_j)$ （添字 i のみについて総和）
 5. $p(a_i, b_j) = p(a_i | b_j) p(b_j)$
 $p(a_i | b_j)$ は**条件付き確率**
- 結合エントロピー
$$H(AB) = - \sum_i \sum_j p(a_i, b_j) \log p(a_i, b_j)$$
これは**結合確率の平均情報量**

3.2.15 条件付きエントロピー

- 結合事象系で、事象系Bの**特定の事象** b_j が観測された（=知った）時の事象系Aのエントロピーは、

$$H(A|b_j) = -\sum_i p(a_i|b_j) \log p(a_i|b_j) \quad (\text{添字 } i \text{ について総和})$$

で定義される。事象 b_j を条件とするAの確率分布のエントロピー。

- これをBの全ての事象 b_j について平均をとった $H(A|B)$ を**条件付きエントロピー**と呼ぶ。

$$H(A|B) = \sum_j p(b_j) H(A|b_j) \quad (\text{添字 } j \text{ について総和})$$

$$= -\sum_i \sum_j p(a_i|b_j) p(b_j) \log p(a_i|b_j) \quad (\text{添字 } i, j \text{ について総和})$$

$$= -\sum_i \sum_j p(a_i, b_j) \log p(a_i|b_j) \quad (\text{同上})$$

- 条件付きエントロピーの性質

$$H(AB) = H(A|B) + H(B) = H(B|A) + H(A)$$

『結合エントロピーは、一方の事象系を条件とする条件付エントロピーと、その事象系自身のエントロピーの和に等しい。』

- 条件付きエントロピーは後述の相互情報量の定義において使われる。

3.3 無記憶情報源の拡大(1)

- シンボルを1つずつ扱うより、いくつかまとめたブロックとして扱うほうが良い場合がある。
 - 例： $S = \{0, 1\}$ の出力アルファベットをもつ2元情報源の出力系列を、2シンボルずつ区切って一つのシンボルとすると、各ブロックの値は00, 01, 10, 11の4種類となる。

情報源 S

$\{0, 1\} \rightarrow 0, 1, 1, 0, 0, \dots$

S の拡大情報源 (S^2)

$\{00, 01, 10, 11\} \rightarrow 01, 11, 00, 10, \dots$

- 定義

S を無記憶情報源 $\{s_1, s_2, \dots, s_q\}$, s_i 生起確率を P_i とする。長さ n の s_i のある系列 (q^n 個) をシンボルとする情報源 $\{\sigma_1, \sigma_2, \dots, \sigma_{q^n}\}$ が $P(\sigma_i) = P(s_{i_1}) P(s_{i_2}) \cdots P(s_{i_n})$ で与えられるとき、**情報源 S の n 次拡大 (S^n と表記)** という。

3.3 無記憶情報源の拡大(2)

- S の n 次拡大情報源 (S^n) のエントロピー

$$\begin{aligned}
 H(S^n) &= \sum P(\sigma_i) \log 1/P(\sigma_i) = -\sum P(s_{i1}) \\
 &\quad P(s_{i2}) \cdots P(s_{in}) \log P(s_{i1})P(s_{i2}) \cdots P(s_{in}) = \sum_{(i1)} \sum_{(i2)} \cdots \\
 &\quad \sum_{(in)} P(s_{i1}) P(s_{i2}) \cdots P(s_{in}) \log P(s_{i1})P(s_{i2}) \cdots P(s_{in}) = - \\
 &\quad \sum \sum \cdots \sum P(s_{i1})P(s_{i2}) \cdots P(s_{in}) \log P(s_{i1}) - \cdots = H(S) + \\
 &\quad H(S) + \cdots + H(S) = nH(S)
 \end{aligned}$$

- 解釈

- 拡大情報源はもとの情報源と実質的に同じ。
- 出力シンボルを n 個ずつ区切って眺める点が違うのみ。
- 元情報源の n シンボル = 拡大情報源の 1 シンボル、となるので、1 シンボルあたりのエントロピーは、拡大情報源が元情報源の n 倍になる。

情報源の中間まとめ

- 情報源の分類

- 無記憶（＝記憶をもたない）情報源

- 特に、時間の概念が導入される情報源において、各時点における情報源シンボルの発生が同一の確率分布に従うとき、無記憶定常情報源、ということがある。
 - 即ち、時間軸をずらしても性質が変わらない、ということ。
 - 定常情報源の出力 $X_1, X_2, X_3, \dots, X_n$ の結合確率分布は次のように書ける

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i)$$

- 記憶をもつ情報源

- 記憶を持つ情報源は、一般的にその範囲が広すぎるので性質を調べるには適さない。このため、より制限した特別な情報源を対象とする。例えば次のもの：

- 記憶範囲を制限したマルコフ情報源
 - 時間的な定常性をもつ定常情報源
 - エルゴード情報源

3.4 マルコフ (Markov)情報源

なぜマルコフなのか？

- ・ 無記憶情報源は余りにも単純すぎる。
- ・ 理論的には扱いやすいが実際の情報源とは異なっている。
- ・ 少し複雑だが現実のモデルに近く、しかも数学的に簡単に扱える情報源が望ましい (→マルコフ情報源)

Andrei Andreyevich Markov (1856–1922),
ロシアの数学者 (確率・統計論)

3.4 マルコフ(Markov)情報源

- 通常の事象は独立事象であることは少なく、相関関係がある。従って、相関がある事象を発生する情報源を扱う必要がある

- 定義

情報源シンボル s_i の生起確率が、その前のいくつか（例えば m 個のシンボル）に依存する情報源をマルコフ情報源という。

特に、 $m = 1$ のときを単純マルコフ情報源、 m 個のときを m 重マルコフ情報源という。

- 数学の確率論ではマルコフ連鎖(Markov chain)又はマルコフ過程(Markov process)と呼ばれる

3.4.1 マルコフ情報源の表現

- M重マルコフ情報源は、アルファベット S と条件付き確率 $P(s_j | s_{j-1}, s_{j-2}, \dots, s_{j-m})$ で完全に定まる

ある時点で、どのシンボルがどの確率で発生するかは、 m 個の過去のシンボルを与えれば決まる

- 過去の m 個のシンボルをマルコフ情報源の状態という。シンボル数 $=q$ のとき、状態数 $=q^m$
- 状態遷移図で表現される

3.4.2 マルコフ情報源の一般的な例

- ある言語の文字集合（例：英語、日本語）

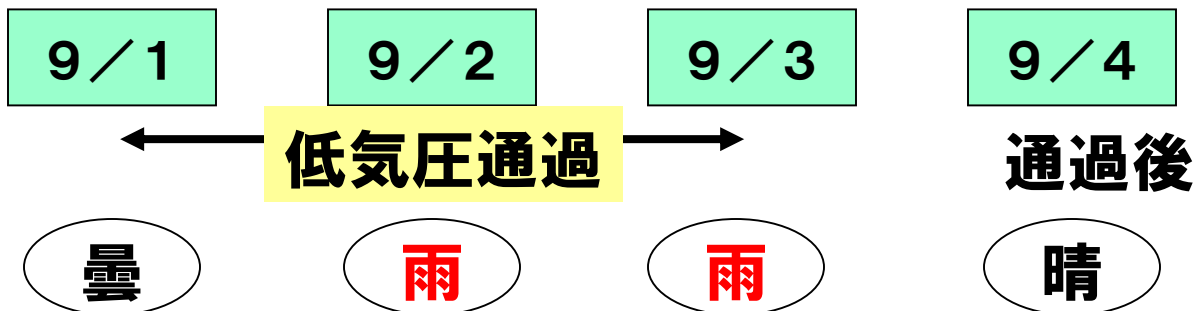
第0次近似（ランダム文字列）→言語になっていない

第1次近似（各文字の確率だけ）→無記憶情報源。言語に程遠い。

第2次近似（2文字までの確率）→マルコフ情報源

第3次、4次、...近似→マルコフ情報源。自然言語に近づく。

- 一般にある場所の天気



前日、前前日の天気に影響されるため、9/1以降の数日は晴より曇や雨の確率の可能性が高い。

3.4.3 マルコフ情報源の例：英語 (1/3)

- 第0次近似モデル（**完全ランダム**）

- アルファベット26文字と空白の27個のシンボルが等確率で生じると考えるモデル。
 - 例えば乱数表から作られる以下の文字列（各文字列は独立で等確率で生起すると仮定）

Xpoml Rxkhrjeffjuj Zlpwcffwkcyj

- 第1次近似モデル（**無記憶情報源**）

- シンボルの生起確率が、実際の英語のシンボル生起確率に等しくして、同じく乱数表を用いて作られるモデル。
 - 文字は独立であるが、英語の頻度分布で生起する

Ocro Hli Rgwr Nmielwis Eu LI Nbnesebya Th Eei

- 自然言語に程遠いが、単語の長さに英語らしい特徴が出始める

- 第2次近似モデル（**単純マルコフ情報源**）

- 1つ前までのシンボルの影響を考慮したモデル。これは単純マルコフ情報源モデル。
 - 条件付確率 $p(s_j | s_i)$, $i, j=1, 2, \dots, 27$ を知ることによって系列を作り出すことができる。

3.4.3 マルコフ情報源の例：英語 (2/3)

- シャノンにより考案された方法

- 代表的と思われる英語の書籍1冊を選ぶ。ランダムに1ページを開き、そのページからランダムに1文字選ぶ。これを「0」とし、最初の文字とする。
- またランダムにページを開き、そのページに書かれた文字を最初から調べて行き、「0」に達するまで進む。達したら、その次の文字（これを「N」とする）を読み取り、第2番目の文字とする。
- 次にランダムにページを開き、同様に文字「N」が見つかるまで進み、見つければその次の文字を第3番目の文字とする。
- 以上の手順を繰り返すことにより、第2近似の系列が作られる。

On ie antsoutinys are inctore st be s deamy achin d ilonasive tucoowe

- On, are, beなどの英語の単語が現れる。

- 第3次近似モデル（2重マルコフ情報源）

- 同様な手法で系列を生成できる。

- 上記と同様に書籍からランダムにページを開き、2重文字（例：In）を選ぶ。
- 別のページを開き、Inの次の文字を読む。これを空白とする。
- 別のページを開き、そのページのn_(空白)の次を読む。これをNとする。
- 以下、これを繰り返す。

- In no lat whey cratict froure birs grocid pondenome of demonstures ... 益々英単語に近くなる

3.4.3 マルコフ情報源の例：英語 (3/3)

- **単語列による第1次近似**

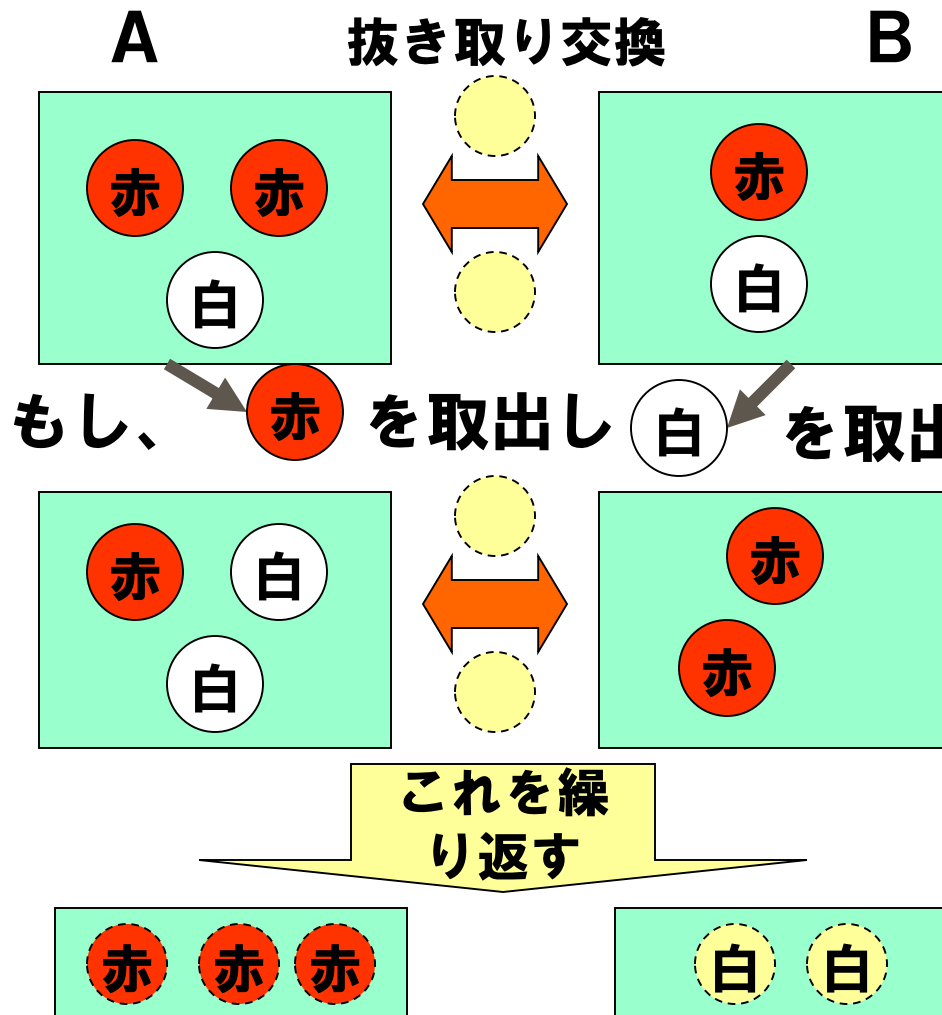
- これまで文字を出力シンボルと考えて情報源を近似してきた。ここでは、出力シンボルが単語と考えて近似を考える。
- (近似的な単語列の作り方)
 - 典型的な書籍を選び、ランダムにあるページを開き、ランダムに適当な単語を選ぶ。(先の例は「文字」、今回は「単語」)
 - 同様の手順を繰り返せば、生起確率が実際の英語と一致する、記憶のない情報源の代表的系列が得られる。
 - Representing and speedily is an good apt or come can different natural here he the a in came the to of to expert

- **単語列による第2次近似**

- 単語を選ぶ時に「1つ前に使われた単語を考慮する」ことにより、単純マルコフ情報源としての単語列による近似が得られる。
 - The head and in frontal attack on an english writer that the character of this point is therefore another method for the letters

3.4.4 数学的に定義されるマルコフ情報源

• 壺のモデル



「ある時点」の玉状態が、
「前」の玉状態に依存して
決まる性質をもつ

最初の状態 (S_1)

1 回交換後の状態 (S_2)

⋮

N 回交換後の状態 (S_n)

3.4.5 マルコフ情報源の表現

$S = \{0, 1\}$ (2元アルファベット) をもつ2重マルコフ情報源

- **遷移確率行列による表現**

- 条件付き確率が、次で与えられる

$$p(0|00) = p(1|11) = 0.8$$

$$p(1|00) = p(0|11) = 0.2$$

$$p(0|01) = p(0|10) = p(1|01) = p(1|10) = 0.5$$

- **状態遷移図 (シャノン線図) による表現**

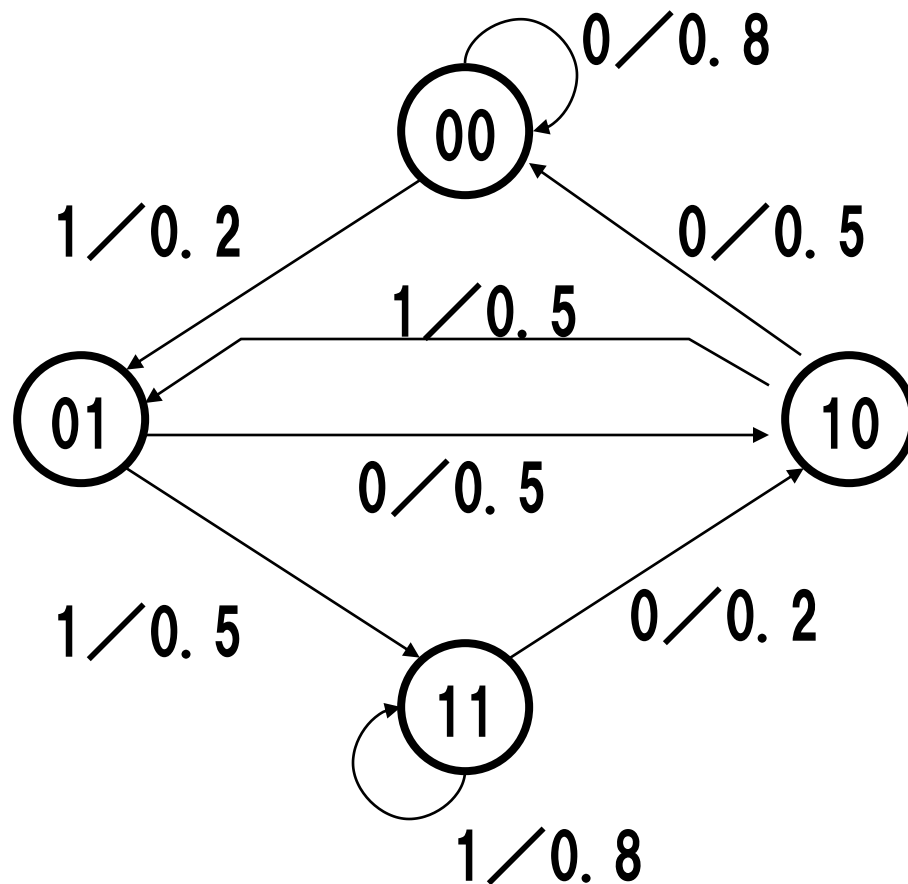
- 前の2つの系列00, 01, 10, 11の条件付き確率により次のシンボルの確率が決まる
- 001101001100は、状態00から、01,11,10,01, ...と遷移する系列として得られる

3.4.6 マルコフ情報源の遷移確率の例

		状態（後）			
		00	01	10	11
P = 状態 （前）	00	0.8	0.2	0	0
	01	0	0	0.5	0.5
	10	0.5	0.5	0	0
	11	0	0	0.2	0.8

- 行方向、列方向に遷移前後の状態を並べる
- 状態遷移の確率を交点に記す

3.4.7 マルコフ情報源のシャノン線図の例



3.4.8 マルコフ情報源の表現：一般化

- **M重マルコフ情報源**は、アルファベット S と条件付き確率 $P(s_j | s_{j-1}, s_{j-2}, \dots, s_{j-m})$ で完全に定まる

ある時点で、どのシンボルがどの確率で発生するかは、**m個のさかのぼったシンボルを与えれば決まる**

- さかのぼった m 個のシンボルを**マルコフ情報源の状態**という。シンボル数 $= q$ のとき、状態数 $= q^m$
- 状態の遷移（移り変わり）図で表現される

3.4.9 遷移確率行列の一般化

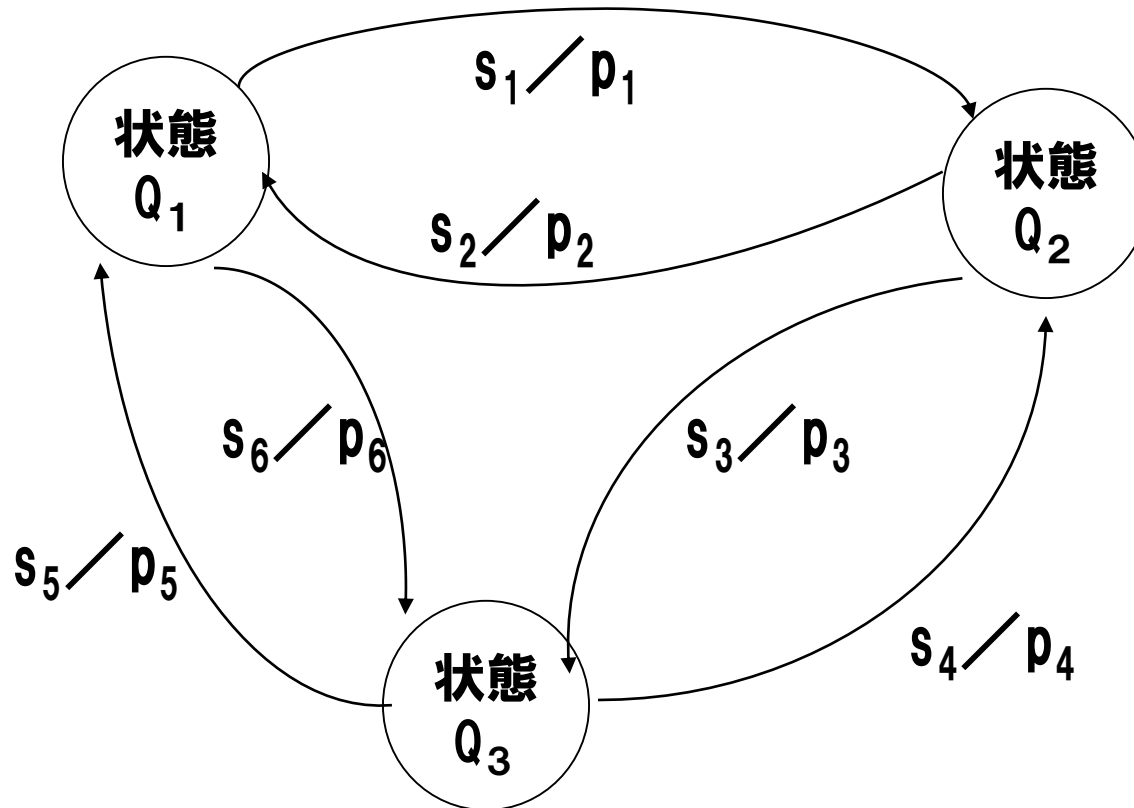
- マルコフ情報源の記号系列がその前の m 個の系列 $s_1, s_2, \dots, s_m$ で決まるとき、可能な s_j の系列を状態とし、次の記号 s_j を状態遷移として生じる行列で表すことができる
- これを遷移確率行列という

$$P = \begin{pmatrix} p_{11}, & p_{12}, & p_{13}, & \dots, & p_{1n} \\ p_{21}, & p_{22}, & p_{23}, & \dots, & p_{2n} \\ \dots & & & & \\ p_{n1}, & p_{n2}, & p_{n3}, & \dots, & p_{nn} \end{pmatrix} \quad \text{行確率の和} = 1$$

$$p_{jk} = p(q_k = s_{j2} \cdots s_k \mid q_j = s_{j1} s_{j2} \cdots s_{jm})$$

n : 状態数

3.4.10 シヤノン線図の一般化



- 各状態を円で表し、状態から状態への遷移を矢線で表す。
線上に遷移確率（および出現記号）を書く
- 状態数： 単純マルコフ=2、2重マルコフ=4、...

3.4.11 高次の推移確率

- 状態 q_i から出発し n 回の遷移で状態 q_j に到達する確率 $p^{(n)}_{ij}$ で表すと、

$$p^{(1)}_{ij} = p_{ij}$$

$$p^{(2)}_{ij} = \sum_v p_{iv} p_{vj}$$

...

$$p^{(n+1)}_{ij} = \sum_v p_{iv} p^{(n)}_{vj}$$

- 遷移行列で表現すると行列の積となる

$$P^{(1)} = P, \quad P^{(2)} = P^2, \quad \dots$$

3.4.12 2 次の推移確率

- $p^{(2)}_{ij} = p_{i1} p_{1j} + p_{i2} p_{2j} + \dots + p_{in} p_{nj}$ 例えば具体的に書けば、

$$p^{(2)}_{11} = p_{11} p_{11} + p_{12} p_{21} + p_{13} p_{31} + \dots + p_{1n} p_{n1}$$

$$p^{(2)}_{12} = p_{11} p_{12} + p_{12} p_{22} + p_{13} p_{32} + \dots + p_{1n} p_{n2}$$

$$p^{(2)}_{13} = p_{11} p_{13} + p_{12} p_{23} + p_{13} p_{33} + \dots + p_{1n} p_{n3}$$

$$P = \begin{matrix} q1 \\ q2 \\ q3 \\ q4 \end{matrix} \begin{pmatrix} q1 & q2 & q3 & q4 \\ 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{pmatrix}$$

$$P^2 = \begin{matrix} q1 \\ q2 \\ q3 \\ q4 \end{matrix} \begin{pmatrix} q1 & q2 & q3 & q4 \\ 0.64 & 0.16 & 0.1 & 0.1 \\ 0.25 & 0.25 & 0.1 & 0.4 \\ 0.4 & 0.1 & 0.25 & 0.25 \\ 0.1 & 0.1 & 0.16 & 0.64 \end{pmatrix}$$

付録

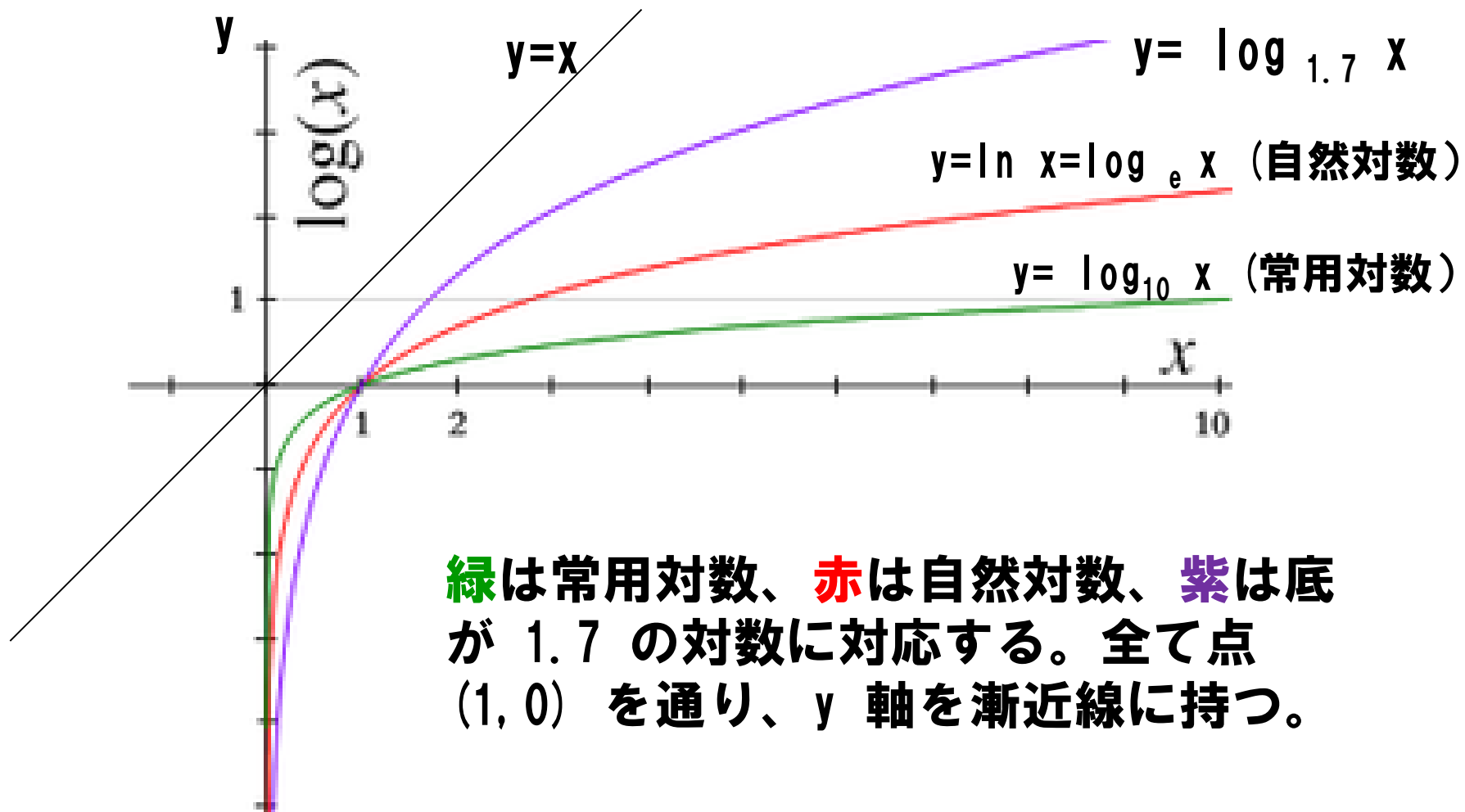


付録A 対数

付録B 確率論

付録C 線形代数：行列論

A.1 対数関数のグラフ



緑は常用対数、**赤**は自然対数、**紫**は底が 1.7 の対数に対応する。全て点 $(1, 0)$ を通り、 y 軸を漸近線に持つ。

出典：Wikipediaより一部複製

A.2 対数の基本的な演算

$$a^{\log_a x} = x$$

$$\log_a (xy) = \log_a (x) + \log_a (y) \quad (\text{積の対数})$$

$$\log_a x^p = p \log_a x \quad (\text{累乗の対数})$$

$$\log_a \frac{1}{x} = \log_a x^{-1} = -\log_a x \quad (\text{逆数の対数})$$

$$\log_a \frac{x}{y} = \log_a x \cdot y^{-1} = \log_a x - \log_a y \quad (\text{分数の対数})$$

$$\log_a x = \frac{\log_b x}{\log_b a} \quad (\text{底の変換公式})$$

A.3 自然対数 (ln)

- ネイピア数 (e)を底とする対数
- 通常のlogと区別するため、lnと表記
 - lnは、natural logarithm、の略から来ている
- 自然対数で定める関数、ln xは、指数関数、 e^x 、の逆関数である
- eの定義としては、次のようなものがよく使われる

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

- $e=2.718281....$ (無限小数、超越数)

B.1 確率の基礎(1)

- 事象、標本空間

- 例えば硬貨投げの実験を1回行った時に起こる結果は、表か裏かの2通り。これらを表=H(ead), 裏=T(ail)で表すと、同じ実験を2回続けて行えば結果は
- 表表 (HH), 表裏 (HT), 裏表 (TH), TT (裏裏), の4通り

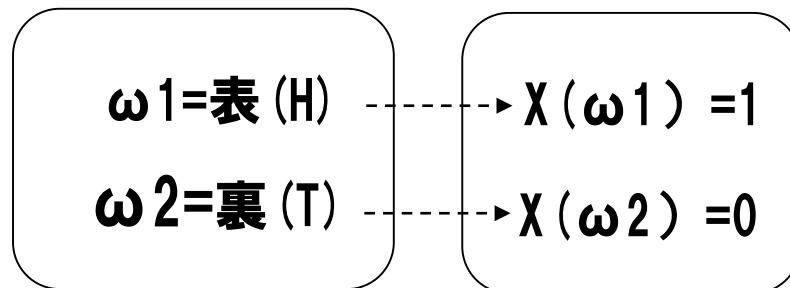
- 用語の定義, 意味

- ある結果を得るのに必要な操作を**試行**という
- 試行によって得られる結果を**事象**という.
- 1つの実験について起こりえる全ての異なる結果の集まりを,**標本空間**という. それぞれの事象のことを標本点と呼ぶ.
- 標本空間で個々の事象にその生起する確率が与えられているものを**確率空間**という.
 - 例: 試行=硬貨投げの場合, 表の出る確率は $1/2$, 裏の出る確率も $1/2$ で与えられるので, 標本空間=確率空間= ({H,T} , { $1/2, 1/2$ })

B.1 確率の基礎(2)

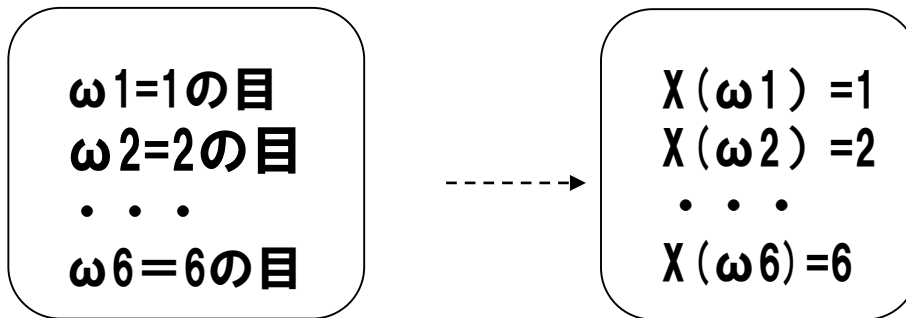
- 事象の結果にある数値を対応させることがある。
（例）硬貨投げの結果，表(H)に対して1を，裏(T)に対して0を対応させる．このように，ある規則に従い，数値を対応させることは関数を定義することに等しい．
- 事象の関数を**確率変数**と呼ぶ．

例 1 : 硬貨投げの標本空間 (Ω)，標本点 ($\omega 1$, $\omega 2$)，確率変数 $X(\omega)$



B.1 確率の基礎(3)

例2：サイコロ投げの標本空間 (Ω)， 標本点 (ω_1, ω_2)， 確率変数 $X(\omega) = X$



- 確率変数の取る値(x)に対して対応づけた実数値 $P(x)$ = 生起確率

例) サイコロ投げの場合, $P(x=1)=P(x=2)=\cdot \cdot \cdot =1/6$

- (確率空間の) **積空間**
 - 2種類の試行があり, それぞれに確率空間 X, Y が, $X=\{x_1, x_2, \dots\}$, $Y=\{y_1, y_2, \dots\}$ と定義されているとする. X の任意の要素と Y の任意の要素の対から成る新しい事象 xy の集合を, **積空間** XY という.

B.2 同時確率，周辺確率，条件付き確率

- 同時確率，周辺確率，条件付き確率

- 積空間 XY に属する事象 $x_i y_j$ の確率を， $P(x=x_i, y=y_j)$ または $P(xy=x_i y_j)$ と書いて，**同時確率**という。
- 同時確率から一方の確率変数の特定事象の確率が，他方の確率変数の事象の確率和として求められる。これを，**周辺確率**という。

$P(x=x_i) = \sum_j P(x=x_i, y=y_j)$ ，総和は変数 y の値に対してとる

- 同時確率を周辺確率で割った値を**条件付き確率**という。

$$P(y=y_j | x=x_i) = P(x=x_i, y=y_j) / P(x=x_i)$$

或いは簡略表現で表せば， $P(y|x) = P(x,y)/P(x)$

事象 x_i が生起した（＝観測された）という条件のもとで，事象 y_j が生起する（＝観測される）確率，を意味する。（制約条件のほうを" $|$ "の右側に書く）

- 条件付き確率を，**事後確率**ともいう。これに対して，条件が付かない確率，上記表現では， $P(y)$ を**事前確率**という。

B.3 条件付確率の表現法

- $P(0|01)$ という表記で、事象01が起こったという条件でさらに事象0が起こる確率Pを表す。
- 条件付確率の記法
 - 縦棒（|）の後に書く値＝条件となる事象
 - 縦棒の前に書く値＝発生事象
- $\sum_i P(a_i | b_j) = 1$ が成立する
 - これは、どのような条件事象 b_j の場合でも、 a_i の起こる確率を全ての a_i について合計すると、どれかは必ず起こるので1になる、ことを表す。

B.4 確率的独立性

- 確率的独立性

- 事象 x_i, y_j の同時確率と個々の周辺確率との間に, 次の関係が成り立つとき, x_i と y_j は確率的に独立という.
- $p(x_i, y_j) = P(x_i) \cdot P(y_j)$
- 全ての i, j の組について成り立てば, 空間 X と Y は確率的に独立であるという.
- 確率的に独立性は, 条件付き確率を使って表せば,
- $P(y=y_j | x=x_i) = P(x=x_i, y=y_j) / P(x=x_i)$
 $= P(x=x_i)P(y=y_j) / P(x=x_i) = P(y=y_j)$ と書ける.
- あるいは, $P(y | x) = P(y)$, と簡易表現することもできる.

B.5 ベイズ (Bays)の定理

- ベイズ (Bays)の定理

$$P(y=y_j | x=x_i) = P(x=x_i, y=y_j) / P(x=x_i) \quad \cdots (1)$$

条件付き確率の関係式を下記のように簡略表現する：

$$P(y_j | x_i) = P(x_i, y_j) / P(x_i) \quad \cdots (2)$$

(2) 右辺の分子，分母は次の式で書き換えられる。

$$\text{分子} = P(x_i | y_j) \cdot P(y_j)$$

$$\text{分子} = \sum_j P(x_i, y_j) = \sum_j P(x_i | y_j) \cdot P(y_j)$$

これらを式 (2) 右辺に代入すると，

$$P(y_j | x_i) = P(x_i | y_j) \cdot P(y_j) / \sum_j P(x_i | y_j) \cdot P(y_j) \quad \cdots (3)$$

式 (3) をベイズの定理という。

これは条件付き確率を条件となる変数を交換して表現したもので，情報理論では雑音のある通信路における復号誤り率の計算，他で使われる重要な式，

B.6 完全事象系

- 完全事象系

- 互いに素（排他的）な事象 $E_1, E_2, \dots, E_n$ があり，それらの確率を $p_i = P(E_i)$, $i=1, 2, \dots, n$ とする．
- 事象系が E_i ですべて覆われ，かつ個々の E_i, E_j 間は独立とする．即ち，

$$\sum_i E_i = E \text{ (全体)} \quad \dots\dots (1)$$

$$\sum_i p_i = 1 \quad \dots\dots (2)$$

$$p(E_i, E_j) = 0 \quad \dots\dots (3)$$

の3つが成り立つとき，系 E は**完全事象系**という。

$$E = \begin{pmatrix} E_1 & E_2 & E_3 & \cdot & \cdot & \cdot & E_n \\ p_1 & p_2 & p_3 & \cdot & \cdot & \cdot & p_n \end{pmatrix}$$

C.1 線形代数_行列(1)

- 行列

- ある集合の要素を矩形に並べたもの

- ある集合とは、一般的にいえば環をなす集合であるが、代表的には整数、実数、有理数、複素数などの数が該当する。

- 行数 n ，列数 m は異なってよい． $n=m$ のとき正方行列という．

- $$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdot & \cdot & \cdot & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdot & \cdot & \cdot & a_{2m} \\ a_{31} & a_{32} & a_{33} & \cdot & \cdot & \cdot & a_{3m} \\ & \cdot & \cdot & \cdot & & & \\ a_{n1} & a_{n2} & a_{n3} & \cdot & \cdot & \cdot & a_{nm} \end{pmatrix}$$

C.2 線形代数_行列(2)

- 行列の和， 差， 積

- 行列の和は， 各要素ごとの和
- 行列の差は， 各要素ごとの差
- 行列の積は， 次のように定義される：

l 行 m 列の $A=(a_{lm})$ ， m 行 n 列の $B=(b_{mn})$ の積行列 $C=(c_{ln})$ は，
 l 行 n 列行列となり， 要素は次式で計算：

$$c_{ij} = \sum_k a_{ik} \cdot b_{kj}, \quad (k=1, \dots, m)$$

積が可能になるには， A の列数（第2添字の最大数）
と B の行数（第1添字の最大数）は同じでなければならない。

- 行列の積は交換律が成り立たない

$$AB \neq BA$$