# 情報理論 配布資料 #2

担当：井上 純一 (情報科学研究科棟 8-13)
URL：http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/

平成 17 年 4 月 25 日

## 目 次



**演習問題 1 の解答例**

(1) 遷移行列の定義から直ちに

$$\boldsymbol{P} = \begin{pmatrix} P(0|0) & P(0|1) \\ P(1|0) & P(1|1) \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{pmatrix} \tag{6}$$

となる．

(2) 入力の状態ベクトルを $\boldsymbol{x}^t = (P(0), P(1))$, 出力の状態ベクトルを $\boldsymbol{y}^t = (P_0, P_1)$ とすれば (肩の添え字の $t$ は転置を表すものと約束する), これらのベクトルと遷移行列の間の関係は

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x} \tag{7}$$

のように書ける．これはまた, 成分で書き直せば

$$P_0 = P(0|0)P(0) + P(0|1)P(1) = \sum_{i=0,1} P(0|i)P(i) \tag{8}$$

$$P_1 = P(1|0)P(0) + P(1|1)P(1) = \sum_{i=0,1} P(1|i)P(i) \tag{9}$$

となり, 既に見た, 2 元通信路 (この場合「対称」ではないことに注意 !) の出力確率の式と一致する．従って, 遷移行列に入力の状態ベクトルを作用させることによって出力の状態ベクトルを算出することができる．ところで, 1 を入力したのであれば, 入力の状態ベクトルは $\boldsymbol{x}^t = (0, 1)$ であるから, これを小問 (1) で求めた遷移行列に作用させれば $P_0 = P_1 = 1/2$ が得られる．

(3) $\boldsymbol{P}^2$ を計算すると

$$\boldsymbol{P}^2 = \begin{pmatrix} \frac{11}{18} & \frac{7}{12} \\ \frac{7}{18} & \frac{5}{12} \end{pmatrix} \tag{10}$$

であるから, この遷移行列で与えられる通信路をグラフ表現すると図 6 のようになる.

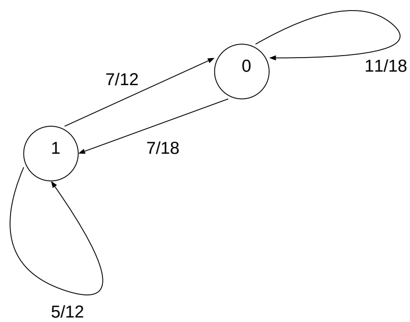


図 6: 遷移行列が $\boldsymbol{P}^2$ で与えられる通信路のグラフ表現.

## 2 エントロピー

前回の講義では, ある事象 $E$ が生起する確率を $p(E) = p$ としたとき, その事象が生じたことがわかったときに我々が得る情報量を $-\log_2 p$ であるとして定義した. ところで, 1 つのコインを投げたとき, 裏, 表の出る確率はどちらも 1/2 であり, 「裏が出た」あるいは「表が出た」という通報で得られる情報量はともに $-\log_2(1/2) = 1$(ビット) である. しかし, この手の通報を実際に受ける前に「どの程度の情報量が得られると期待できるか ?」ということを考えてみることは意味があるであろう. このコイン投げの例で言えば, 裏である確率が 1/2 であるから $-\log_2(1/2)$ もの情報量を 1/2 の確率で, また, 表である確率も 1/2 であるから, $-\log_2(1/2)$ もの情報量を 1/2 で得ることになる. 従って, 得られる情報量の期待値 $H$ は

$$\begin{aligned} H &= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1(\text{ビット}) \\ &= -p(\text{裏})\log_2 p(\text{裏}) - p(\text{表})\log_2 p(\text{表}) \\ &= -\sum_{i=\text{裏, 表}} p(i)\log_2 p(i) \end{aligned} \tag{11}$$

と書け, 我々が受け取るであろう期待値という意味での情報量もやはり 1(ビット) であることがわかる. このようにして算出される情報量の期待値を**平均情報量**あるいは**エントロピー**と呼んでいる[1] .

これらをふまえると, エントロピーは一般に次のように定義される.

**エントロピー**

ある事象の集合を $\mathcal{A} = \{a_1, a_2, \cdots, a_m\}$ とし, $P_X(x)$ を $x \in \mathcal{A}$ という事象が生起される確率とするならば, 確率変数 $X$ のエントロピーは

$$H(X) = -\sum_{x\in\mathcal{A}} P_X(x)\log_2 P_X(x)$$

で与えられる.

後に見るように, ある $x \in \mathcal{A}$ という特定の事象が確率 $P_X(x) = 1$ で確実に起こるならば, $H(x) = 0$ とな

[1] 前回学んだ $-\log_2 p$ を「自己情報量」と呼ぶ場合もある.

り, それ以外では一般に $H(x) > 0$ である. 従って, エントロピーは「確率変数 $X$ に関する我々の知識のあいまいさ」と考えることができる.

具体的にはこの定義式から 2 枚のコインを投げで表の出た回数を確率変数 $X$ としたとき, $X$ のエントロピー等が計算できる (教科書 p. 15, **例 2・2** 参照).

## 2.1 エントロピーの性質

上に定義したエントロピーには次のような性質がある.

(1) **$H(x) \geq 0$ である.**
$0 \leq P(x) \leq 1$ とこの区間での対数関数 $\log_2 P(x)$ の性質から明らか.

(2) **$P(x) = 1$ となる $x \in \mathcal{A}$ があれば, $H(x) = 0$ である.**
どれか 1 つの事象 $x$ が確実に起きる $P(x) = 1$ であることがわかれば, 確率変数 $X$ についてのあいまいさはゼロである.

(3) **$0 < P(x) < 1$ となる $x \in \mathcal{A}$ があれば $H(x) > 0$ である.**
$-P(x) \log P(x) > 0$ と $H(x) \geq 0$ より明らか.

## 2.2 2 値エントロピー関数

$P(a) = p, P(b) = 1 - p$, $\mathcal{A} \in \{a, b\}$ とするとき, $p$ の関数となるエントロピー $h(p)$ は

$$\begin{aligned} h(p) &= -p(a) \log_2 p(a) - p(b) \log_2 p(b) \\ &= -p \log_2 p - (1-p) \log_2 (1-p) \end{aligned} \tag{12}$$

と書ける. この $p$ の関数 $h(p)$ を **2 値エントロピー関数**と呼ぶ. 図 7 にこの関数を図示する. この図 7 より,

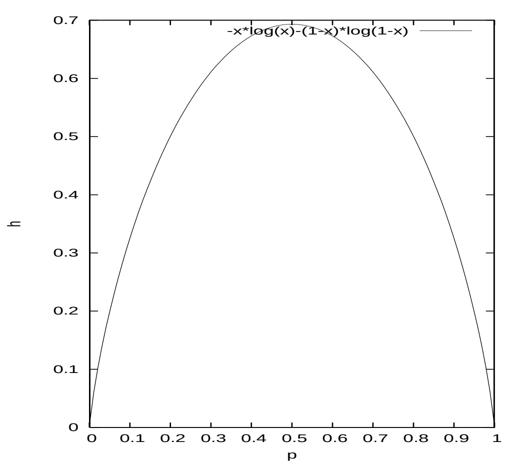


図 7: 2 値エントロピー関数 $h(p)$.

2 値エントロピー関数は $p = 1/2$ で最大値 $h(1/2) = 1$ をとる (図では自然対数を用いていることに注意) ことがわかる. この 2 値エントロピー関数を一般化する問題を例題として見ておこう.

**例題 2**

2 値エントロピー関数：$H(p) = -p\log p - (1-p)\log(1-p)$ の最大値が 1 であり, そのときの $p$ の値が $p = 1/2$ であることを学んだ. ここでは, これを $n$ 個の事象に拡張することを考えたい. つまり, 事象 $1, 2, \cdots, n$ の生起確率を $p_1, p_2, \cdots, p_n$ とするときの $n$ **値エントロピー関数**：

$$H = -\sum_{i=1}^{n} p_i \log p_i$$

の最大値を与える $p_i\ \ (i = 1, \cdots, n)$ の値と, そのときの最大値を考えよう. 以下の問いに答えよ.

(1) $\lambda$ を定数とするとき, 次に定義される $n$ 変数関数 $F(p_1, p_2, \cdots, p_n)$：

$$F(p_1, p_2, \cdots, p_n) = H + \lambda\left(1 - \sum_{i=1}^{n} p_i\right)$$

に対し, $F$ を最大とするような $p_k^*\ \ (k = 1, \cdots, n)$ は

$$p_k^* = \mathcal{P}(\lambda)\ \ (k = 1, \cdots, n)$$

であることを示し, $\lambda$ の関数 $\mathcal{P}$ を求めよ.

(2) (1) の結果と $n$ 個の事象の生起確率の和は 1 であるという条件：

$$\sum_{k=1}^{n} p_k = 1$$

を用いて $\lambda$, 及び, $p_k^*\ \ (k = 1, \cdots, n)$ を $n$ を用いて表し, $H^* = H(p_1^*, p_2^*, \cdots, p_k^*)$ を求めよ.

このとき, $H^*$ が求める $n$ 値エントロピー関数の最大値であり, $p_k^*\ \ (k = 1, \cdots, n)$ がその最大値を与える $p_k\ \ (k = 1, \cdots, n)$ となる[a].

[a] ここでの例のように, $\sum_i p_i = 1$ 等の制約条件下で関数 $H(\{p_i\})$ の最大値を求める方法を**ラグランジュの未定係数法**と呼び, $\lambda$ は**ラグランジュの未定係数**と呼ばれる.

**(解答例)**

ラグランジュの未定係数法を用いることにより, $n$ 値エントロピー関数の最大値とその最大値を与える確率を求めていく. 問題文の誘導に従って行けばよい.

(1)

$$F(\{p_k\}) = H + \lambda\left(1 - \sum_{i=1}^{n} p_i\right) = -\sum_{i=1}^{n} p_i \log p_i - \lambda \sum_{i=1}^{n} p_i + \lambda \tag{13}$$

に対して, $F(\{p_k\})$ の極値条件より

$$\frac{\partial F}{\partial p_k} = -p_k \cdot \frac{1}{p_k} - \log p_k - \lambda = 0 \tag{14}$$

従って, この方程式の解である $p_k^*$ は

$$p_k^* = 2^{-(1+\lambda)} \equiv \mathcal{P}(\lambda) \tag{15}$$

となる．なお，

$$\left.\frac{\partial^2 F}{\partial p_k \partial p_l}\right|_{p_k=p_k^*, p_l=p_l^*} = 0, \quad \left.\frac{\partial^2 F}{\partial p_k^2}\right|_{p_k=p_k^*} = -\frac{1}{p_k^*} = -2^{(1+\lambda)} < 0 \quad (\forall_{k,l\neq k}) \tag{16}$$

であるから，$p_k^*$ は $F$ の最大値を与える．

(2) 全ての $k$ について，$p_k^*$ を足しあげたものは 1 にならなければならないので

$$\sum_{k=1}^{n} p_k^* = 2^{-(1+\lambda)} \sum_{k=1}^{n} 1 = n \cdot 2^{-(1+\lambda)} = 1 \tag{17}$$

が成り立つべきであり，これを $p_k^*$, 及び, $\lambda$ に関して解けば

$$p_k^* = \frac{1}{n}, \quad \lambda = \log n - 1 \tag{18}$$

が得られる．従って，これらを $n$ 値エントロピー関数 $H$ に代入すれば，$H$ の最大値は

$$H(\{p_k^*\}) = -\sum_{k=1}^{n} p_k^* \log p_k^* = -\frac{1}{n} \times n \log \frac{1}{n} = \log n \tag{19}$$

となる．ここで，$n=2$ とおけば教科書の 2 値エントロピー関数の場合と一致することに注意．

**(参考までに別解)**

この問題では $n$ 値エントロピー関数を最大にする確率が，$n$ 個全ての事象が等確率で現れるとき，すなわち $p_1^* = p_2^* = \cdots = p_n^* = 1/n$ で与えられることは直観的にすぐわかることですから，この解 $p_1^* = p_2^* = \cdots = p_n^* = 1/n$ が実際に $n$ 値エントロピー関数 $H(\{p_k\})$ を最大にすることを「示す」ことによって，ラグランジュの未定係数法を用いることなく解答を作ることもできます．ただし，ここでも確率の規格化条件：

$$\sum_{j=1}^{n} p_j = 1 \tag{20}$$

あるいは同じことですが

$$p_n = 1 - \sum_{k=1}^{n-1} p_k \tag{21}$$

が成り立っていることを忘れないようにしなければなりません．

さて，ここで具体的に示すべきことは次の 2 点です．

- $n$ 値エントロピー関数 $H$ の $p_k$ $(k=1,\cdots,n)$ に関する 1 階微分係数が $p_k^* = 1/n$ でゼロとなる．
- $H$ の $p_k$ に関する 2 階微分係数の作る $n \times n$ の行列式の $p_k = p_k^*$ $(k=1,\cdots,n)$ での値：

$$I^{(n)} = \begin{vmatrix} I_{11} & I_{12} & \cdots & I_{1n} \\ I_{21} & I_{22} & \cdots & I_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ I_{n1} & I_{n2} & \cdots & I_{nn} \end{vmatrix} \tag{22}$$

が行列式のサイズを 1 つずつ増加させるにつれ，その符号を変えて行く．言い方をかえれば，$n$ 以下の任意の自然数 $m$ に対し，$I^{(m)}$ は $(-1)^m$ のような因子を持つ．ここで各要素 $I_{kl}$ は $I_{kl} = (\partial^2 H/\partial p_k \partial p_l)$ の $p_k^* = p_l^* = 1/n$ での値である．

まずは 1 番目の条件ですが, 条件 (21) があるので

$$\frac{\partial p_n}{\partial p_k} = -1 \tag{23}$$

が成り立つことに注意し, $n$ 値エントロピー関数を

$$\begin{aligned} H &= -\sum_{k=1}^{n} p_k \log p_k \\ &= -\sum_{j=1}^{n-1} p_j \log p_j - p_n \log p_n \end{aligned} \tag{24}$$

と書き直しておけば

$$\begin{aligned} \frac{\partial H}{\partial p_k} &= -1 - \log p_k - \frac{\partial}{\partial p_n}\{p_n \log p_n\}\left(\frac{\partial p_n}{\partial p_k}\right) \\ &= \log\left(\frac{p_n}{p_k}\right) \end{aligned} \tag{25}$$

つまり, $k=1,\cdots,n-1$ に対し, $p_k = p_n$, すなわち, $p_1^* = p_2^* = \cdots = p_n^* = 1/n$ で $H$ が極値をとることが示せます.

次は 2 番目の条件, つまり, $p_1^* = p_2^* = \cdots = p_n^* = 1/n$ で $H$ が最大値をとることの証明ですが, 実際それをここで示す前に, この条件の成立は, 本問題でラグランジュの未定係数法を用いる際に導入した関数 $F$ の場合には自明であったことに注意しておきましょう. なぜならば, 関数 $F$ には確率の規格化条件 (21) が既に取り込んであるので, $F$ の中で変数 $p_1,\cdots,p_n$ は全て独立となり, $F$ の 2 階微分からなる行列の非対角要素 $(\partial^2 F/\partial p_k \partial p_l)$ $(k \neq l)$ は全てゼロであり, $m$ 次の行列式は $m$ 個の対角要素を全て掛け合わせればよいだけで, 確かに $\{-2\lambda^{1+\lambda}\}^m \sim (-1)^m$ の因子を持つことが直ちにわかるからです. 一方, ここで考える $n$ 値エントロピー関数 $H$ の場合, これはさほど自明ではありません. というのも

$$\begin{aligned} I_{kl} &= \frac{\partial^2 H}{\partial p_k \partial p_l} \\ &= -\frac{1}{p_k}\left(\frac{\partial p_k}{\partial p_l}\right) + \frac{1}{p_n}\left(\frac{p_n}{\partial p_l}\right) \\ &= \begin{cases} -2n & (k=l) \\ -n & (k \neq l) \end{cases} \end{aligned} \tag{26}$$

となり, 非対角要素が残るからです. 従って, 前出 2 番目の条件を示すために我々が考えるべき行列式は

$$I^{(n)} = \begin{vmatrix} -2n & -n & \cdots & -n \\ -n & -2n & \cdots & -n \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ -n & \cdots & \cdots & -2n \end{vmatrix} \tag{27}$$

であり, $m \leq n$ なる $m$ に対し, $I^{(m)}$ を計算し, これが $(-1)^m$ なる因子を持つことを示すことがここでの目標となります.

これは一見すると厄介そうですが, 行列式の基本変形と余因子展開を用いて次のように変形すれば簡

単な漸化式が得られます.

$$I^{(m)} = \begin{vmatrix} -n & -n & \cdots & -n & -n \\ n & -2n & -n & \cdots & -n \\ 0 & -n & \cdots & \cdots & -n \\ \cdots & \cdots & \cdots & \cdots & -n \\ \cdots & \cdots & \cdots & \cdots & -n \\ 0 & -n & \cdots & \cdots & -2n \end{vmatrix} = (-n)I^{(m-1)} + (-n)K^{(m-1)} \tag{28}$$

ここで行列式 $K^{(m-1)}$ は

$$K^{(m-1)} = \begin{vmatrix} -n & -n & \cdots & -n \\ -n & -2n & \cdots & -n \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ -n & -n & \cdots & -2n \end{vmatrix} = (-n)K^{(m-2)} \tag{29}$$

で定義されます. 従って我々は次の連立漸化式：

$$I^{(m)} = (-n)I^{(m-1)} + (-n)K^{(m-1)} \tag{30}$$

$$K^{(m-1)} = (-n)K^{(m-2)} \tag{31}$$

を解けばよいことになったわけです. (31) は直ちに

$$K^{(m-1)} = (-n)^2 K^{(m-3)} = \cdots = (-n)^{(m-2)} K^{(1)} \tag{32}$$

となり, これを用いれば (30) も同様にして, 逐次的に

$$\begin{aligned} I^{(m)} &= (-n)I^{(m-1)} + (-n)^{m-1}K^{(1)} \\ &= (-n)\{-nI^{(m-2)} + (-n)^{m-2}K^{(1)}\} + (-n)^{m-1}K^{(1)} \\ &= (-n)^2 I^{(m-2)} + 2(-n)^{m-1}K^{(1)} \\ &= \cdots = (-n)^{(m-1)}\{I^{(1)} + (m-1)K^{(1)}\} \end{aligned} \tag{33}$$

と変形できるので, $I^{(1)} = K^{(1)} = -2n$ に注意すれば

$$I^{(m)} = m\,2^{m-1}n^m \underline{(-1)^m} \tag{34}$$

が得られ, 題意が示せました.

これにたどり着くまでに面倒な手続きが必要でしたが, もしもラグランジュの未定係数法というものを知らなければ, このような手順を踏まなければならないということであり, このことは逆に, ここで **(別解)** を例として取り上げることにより, ラグランジュの未定係数法がとても有効な方法であることが実感できたのではないかと思います.

## 2.3　複合事象のエントロピー (結合エントロピー)

2 つの異なる事象の集合を $\mathcal{X} = \{a_1, a_2, \cdots, a_n\}, \mathcal{Y} = \{b_1, b_2, \cdots, b_n\}$ とする. (例えば, $a_1$ (晴れ), $a_2$ (雨), ... $b_1$ (気温：摂氏 0〜10 度), $b_2$ (気温：摂氏 10〜20 度), ... などを表すものとして考えればよい). このとき, **複合事象のエントロピー (結合エントロピー)** を次のように定義する.

$$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\log P_{XY}(x,y) \tag{35}$$

これは 2 つの変数 $X, Y$ のあいまいさを表している.

## 2.4　条件付きエントロピー

$X=x$ という条件のもとでのエントロピー：

$$H(Y|X=x) \quad = \quad -\sum_{y\in\mathcal{Y}} P_{Y|X}(y|X=x)\log P_{Y|X}(y|X=x) \tag{36}$$

を考えよう. ここの $P_{Y|X}(y|X=x)$ は $X=x$ という条件下で $Y=y$ となる条件付き確率を表すものとする. このとき, このエントロピー $H(Y|X=x)$ は $X=x$ の値によるので, この確率変数 $x$ について平均化し

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} P_X(x)H(Y|X=x) \\ &= -\sum_{x\in\mathcal{X}} P_X(x)\sum_{y\in\mathcal{Y}} P_{Y|X}(y|X=x)\log P_{Y|X}(y|X=x) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)P_{Y|X}(y|x)\log P_{Y|X}(y|x) \end{aligned} \tag{37}$$

が得られるが, 確率に関する積の公式から $P_{XY}(x,y)=P_X(x)P_{Y|X}(y|x)$ であるから, これを用いると

$$H(Y|X) \quad = \quad -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{XY}(x,y)\log P_{Y|X}(y|x) \tag{38}$$

が得られる. これを**条件付きエントロピー**と呼ぶ. この条件付きエントロピーは $X$ が与えられたという条件のもとで, 確率変数 $Y$ についての情報に関するあいまいさを表す.

**演習問題 2**

1. A 氏, B 氏, C 氏の 3 人が 3 の目しか出ないように仕掛けがされたサイコロの目を予想するゲームを行う. ただし, A 氏, B 氏, C 氏はそれぞれ次の情報を予め持っているものとする.
   - A 氏はサイコロの目は奇数しか出ないことを知っている.
   - B 氏はサイコロの目は 3 と 5 しか出ないことを知っている.
   - C 氏はサイコロの目は 3 しか出ないことを知っている.

   このとき, A 氏,B 氏,C 氏のサイコロの目に関する知識のあいまいさを表すエントロピー $H_A, H_B, H_C$ を求めよ.

2. コイン投げを 2 回行うものとする. ただし, コインには仕掛けがあり, 2 回目には必ず表が出る ように仕組まれているものとする. 1 回目, 2 回目のコインの裏表を表す確率変数をそれぞれ $\mathcal{A}=\{$ 表, 裏 $\}, \mathcal{B}=\{$ 表, 裏 $\}$ としたとき, 条件付エントロピー $H(\mathcal{B}|\mathcal{A})$ を求めよ.

**(注)**：次週 (5/2) は休講とします. 今回のレポート締め切りは **5/9 の講義開始時まで**.