



HOKKAIDO UNIVERSITY

Title	2005年度 情報理論講義ノート
Author(s)	井上, 純一; Inoue, Jun-ichi
Description	http://www005.upp.so-net.ne.jp/j_inoue/index.html http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/
Issue Date	2005-11-18T09:19:52Z
Doc URL	https://hdl.handle.net/2115/772
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	learning object
File Information	InfoTheory05_1.pdf, 第1回講義ノート



情報理論 配布資料 #1

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 18 日

目次

1	イントロダクション	1
1.1	情報量とは何か？	1
1.2	A4 用紙 1 ページ分の情報量を考える	2
1.3	情報通信路	2
1.4	情報源と通信路の確率モデル	3
1.4.1	情報源の確率モデル	3
1.4.2	通信路の確率モデル	4

1 イントロダクション

まずは情報理論で学ぶことがらをざっと簡単に眺めてみる。各項目は追々詳しく確認していくことになる。

1.1 情報量とは何か？

情報理論では「情報」という抽象的概念の「量」を次のように定義する。

情報量 : $-\log_2 p$. ここに $p(E) = p$ はある事象 E が起こる確率.

(例 1) : 明日の天気に関して「晴れ」か「雨」の 2 通りしかなく, $p(\text{晴れ}) = p(\text{雨}) = 1/2$ であり, 明日は「晴れ」であることがわかったときに得られる情報量は $-\log_2 p(\text{晴れ}) = \log_2 2 = 1$ (ビット, Bit) となる.

(例 2) :

$$\begin{cases} \text{犬が人間に噛みつく確率} & : p(\text{犬} \rightarrow \text{人間}) = \frac{1}{2^3} \\ \text{人間が犬に噛みつく確率} & : p(\text{人間} \rightarrow \text{犬}) = \frac{1}{2^{1000}} \end{cases}$$

このとき, 実際に我々が「犬が人間に噛みついた」状況に遭遇した(「犬が人間に噛みついた」という通報を我々が得た)ときに得られる情報量は $-\log_2 p(\text{犬} \rightarrow \text{人間}) = 3$ (ビット) であるが, 逆に「人間が犬に噛みついた」状況に出くわせば $-\log_2 p(\text{人間} \rightarrow \text{犬}) = 1000$ (ビット) もの情報量が得られる. 前者はニュースにもならない, ありふれた日常の一こまに過ぎないが, 後者はめったにあるものではなく, その地域のニュー

ス程度にはなるのかもしれない¹

従って情報量はある通報で我々が驚く度合いを表していると言ってよい。これはその通報の事象が起こる確率が小さく、めったに起きないような通報に対して大きくなる。

注：上のようにして情報量を計算するためには、事象の起こる確率（生起確率）： p が前もって何らかの計測によってわかっていなければならない。人間に噛みついた犬の例で言うのであれば、保健所が過去数年間にわたり全国でそのような事例をチェックしており、「1日あたりにそのようなことが起こりうる回数」として、この確率は算出されたのかもしれない。しかし、情報理論で扱う対象：音声、画像等は前もってその生起確率がわかっているものとみなされ、また、そのようにして理論が構築されている。

1.2 A4用紙1ページ分の情報量を考える

次の図1のようなA4用紙1ページ分の情報量を考えてみよう。この用紙の1マスには記号、アルファベッ

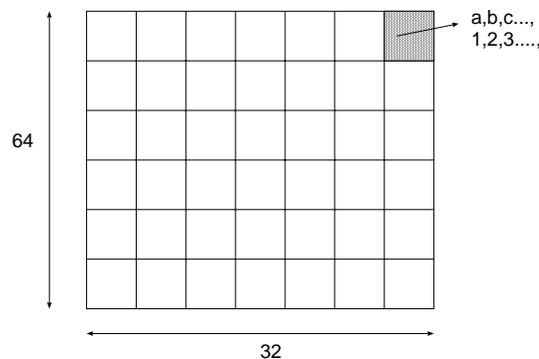


図1: ここで考えるA4用紙1ページは $64 = 2^6$ 行, $32 = 2^5$ 列からなる。

ト、数字等, $2^{12} = 4096$ 種類の対象が書き込めるとする。このとき、ある1マスにある文字が書き込まれる確率は $p = 1/2^{12}$ であるから、ある1マスに入るその文字を見て我々が得る情報量は $-\log_2 p = 12$ (ビット)である。よって、 $2^6 = 64$ 行, $2^5 = 32$ 列からなるA4用紙1ページには $12 \times 2^6 \times 2^5 = 12 \times 2^{11}$ (ビット)もの情報量が含まれていると見積もることができる。ここで、8ビットを1バイト(B), 8×2^{10} ビットを1キロバイト(KB)と定義するので、そのように情報量の単位を約束すれば、このA4用紙1ページの情報量は

$$\frac{12 \times 2^{11}}{8 \times 2^{10}} = 15 \text{ (KB)}$$

ということになる。

1.3 情報通信路

この講義ではアルファベット、記号、数字、漢字等は全て0,1からなる系列であるデジタル情報で表現されるものとする。これと対比されるアナログ情報後に見る標準化定理により、デジタル情報として扱え

¹ もっとも「人間が犬に噛みついた」という通報が社会に対して何らかのインパクトを与えた時代が「かつての日本」にもあるにはあったが、何が起っても不思議ではない昨今ではこの種の情報量ももっと低いかもしれない。つまり、最近では確率： p (人間 → 犬)はもう少し大きくて、我々はこの通報に対し、次のように反応するであろう「人間が犬に噛みついたって？まあ、そういう人もいるだろうね」と。今後は p (犬 → 人間) $<$ p (人間 → 犬) となってしまうのを祈るばかりだ。

る。このような情報を送信するための仮想的な通信路のモデルとして次の 2 種類が知られている。

(a) シャノン (Shannon) の通信路モデル

教科書 p.6 図 1・2 (a) を参照。

(b) ファノ (Fano) の通信路モデル

教科書 p.6 図 1・2 (b) を参照。これは (a) の送信機と受信機の部分を詳細化しており、後にみる「符号化」を詳しく議論する際に便利である。

情報源符号化：情報源のもっている冗長性を取り除くこと (圧縮)。

通信路符号化：通信路に発生する雑音に打ち勝つような符号を構成すること。

誤り訂正符号：雑音の多い通信路では同じ通報を 2 度行う (冗長性を与える) こと。

$$\left\{ \begin{array}{l} \text{冗長性はできるだけ小さな方がよい。} \\ \text{符号化・復号化が簡単な方がよい。} \end{array} \right.$$

1.4 情報源と通信路の確率モデル

1.4.1 情報源の確率モデル

離散的 M 元情報源：決められた時間ごとに M 種類の記号の中から次々と発生させる情報源。

注：情報理論では発生される情報その内容ではなく、情報を構成する記号の順序と、その発生確率で特徴つけられる。

X_n を時刻 n に情報源が発生した記号を表す確率変数とすると、集合： $\{X\} = X_0, X_1, X_2, \dots, X_n$ がその生成される記号系列となり、実際に記号列： $s_0, s_1, s_2, \dots, s_n$ が発生したとすれば、その生成確率は

$$P(X_0 = s_0, X_1 = s_1, X_2 = s_2, \dots, X_n = s_n) = P(s_0, s_1, s_2, \dots, s_n) \quad (1)$$

なる結合 (同時) 分布で特徴つけられる。

(例)：結合分布 $P(x_i, x_j, x_k)$ がわかれば、 x_i の確率分布が与えられたときの x_j, x_k の条件付き確率分布は確率に関する積の公式から

$$P(x_j, x_k | x_i) = \frac{P(x_i, x_j, x_k)}{P(x_i)} \quad (2)$$

のように計算することができる。しかし、記号列のサイズ n が大きくなると、結合分布を求めることが難しくなるので、次のような理想化を用いる。

(i) 記憶のない情報源： $P(x_i, x_j) = P(x_i)P(x_j)$ 。つまり、全ての確率変数 x_i, x_j が独立である。

(ii) 定常情報源： $P(x_i = s) = P(x_j = s) = P(s)$ 。つまり、各時刻において、記号の発生確率が等しい。

定常的単純マルコフ情報源：ある記号 s_i の発生確率が、その直前の記号 s_j のみによって決まり、 $P(s_i | s_j)$

によって与えられる情報源.

例として, s_1, s_2, s_3 の3つの記号 (状態のみ) とすると

$$P = \begin{pmatrix} P(s_1|s_1) & P(s_1|s_2) & P(s_1|s_3) \\ P(s_2|s_1) & P(s_2|s_2) & P(s_2|s_3) \\ P(s_3|s_1) & P(s_3|s_2) & P(s_3|s_3) \end{pmatrix} \quad (3)$$

で定義される行列 P を遷移行列と呼ぶ.

1.4.2 通信路の確率モデル

ここでは

送信記号列: X_0, X_1, \dots, X_n

出力記号系列: Y_0, Y_1, \dots, Y_n

とし, 実際に系列: s_0, s_1, \dots, s_n を送信し, 系列: t_0, t_1, \dots, t_n を受信したとすれば, 条件付き確率分布:

$$P(Y_0 = t_0, Y_1 = t_1, \dots, Y_n = t_n | X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) \quad (4)$$

が通信路の性質を決める. ここでもやはり, 次のような簡略化が行われる.

無記憶定常通信路: 上記の条件つき確率分布が次のように書き直すことができるものとする.

$$P(Y_0 = t_0, Y_1 = t_1, \dots, Y_n = t_n | X_0 = s_0, X_1 = s_1, \dots, X_n = s_n) = \prod_{i=1}^n P(Y_i = t_i | X_i = s_i) \quad (5)$$

(各時刻で入出力関係が独立である).

(例): 2元対称通信路

図2のようにそれぞれの確率分布が $P(0|0) = P(1|1) = 1-p, P(1|0) = P(0|1) = p$ で与えられる. このとき,

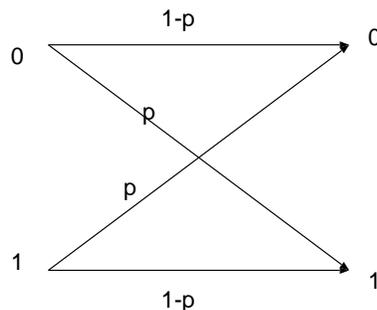


図 2: 2元対称通信路のグラフ表現.

入力情報源の確率モデルとして $0, 1$ がともに等確率で生成されるものを仮定する. つまり, $P(0) = P(1) = 1/2$ とすれば, 出力記号が $0, 1$ をとる確率 P_0, P_1 は次のようにして計算できる.

$$P_0 = \sum_{i=0,1} P(0|i)P(i) = P(0|0)P(0) + P(0|1)P(1) = (1-p) \times \frac{1}{2} + p \times \frac{1}{2} = \frac{1}{2} \quad (6)$$

$$P_1 = \sum_{i=0,1} P(1|i)P(i) = P(1|0)P(0) + P(1|1)P(1) = p \times \frac{1}{2} + (1-p) \times \frac{1}{2} = \frac{1}{2} \quad (7)$$

例題 1

入力信号が 0, 1 であり

$$0 \rightarrow 1, \quad 1 \rightarrow 0$$

という反転がともに確率 p で生じ、また

$$0 \rightarrow x, \quad 1 \rightarrow x$$

のように、入力信号 0, 1 が 0, 1 以外の記号 x へと「消失」してしまう確率を q とするような通信路を考える。このとき以下の問いに答えよ。

- (1) この通信路を教科書 p. 11 の図 1・4 にならってグラフで表現せよ。
- (2) この通信路の特性を条件付確率で表せ。
- (3) 信号 0, 1 がそれぞれ、 $1/2$ の確率で入力されるとするとき、出力信号が 0, 1, x である確率 p_0, p_1, p_x を求めよ。
- (4) この通信路が記憶のない定常通信路であるとする、10 ビットの信号を送信した際、そのうちの k ビットが消失し、残りの $10 - k$ ビットのうち、 l ビットが反転して送信される確率 $p(k, l)$ を求めよ。
- (5) (4) で求めた $p(k, l)$ に対し、次に定義される確率：

$$p_{\text{total}} = \sum_k \sum_l p(k, l)$$

を求めよ。ただし、 $\sum_k \sum_l (\dots)$ は可能な全ての k, l に関する和を表す。ここで、必要ならば

$$(a + b)^n = \sum_{m=0}^n {}_n C_m a^m b^{n-m}$$

なる恒等式を用いてもよい。

(解答例)

(1) 図 3 参照。

(2)

$$p(0|1) = p(1|0) = p, \quad p(x|0) = p(x|1) = q, \quad p(0|0) = p(1|1) = 1 - p - q$$

(3)

$$p_0 = \sum_{k=1,0} p(0|k)p(k) = p(0|1)p(1) + p(0|0)p(0) = \frac{1}{2} \times p + (1 - p - q) \times \frac{1}{2} = \frac{1}{2}(1 - q)$$

$$p_1 = \sum_{k=1,0} p(1|k)p(k) = p(1|1)p(1) + p(1|0)p(0) = (1 - p - q) \times \frac{1}{2} + p \times \frac{1}{2} = \frac{1}{2}(1 - q)$$

$$p_x = \sum_{k=1,0} p(x|k)p(k) = p(x|1)p(1) + p(x|0)p(0) = \frac{1}{2} \times q + \frac{1}{2} \times q = q$$

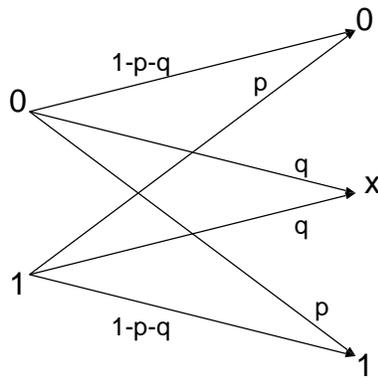


図 3: 例題 1 の通信路に関するグラフ表現.

ここで、出力値としては 0, 1, 及び, x しかないことに注意すれば

$$p_0 + p_1 + p_x = 1$$

が成り立っていることに注意.

- (4) 例えば, $k = 2, l = 3$ の場合には, 2 個の x が, 3 個の b (反転ビット), 5 個の a (正常ビット) からなる 10 個の記号の列:

xxbbbaaaaa
xbaxbbaaaa
...
...

が現れる確率を求めればよい. 上記のような並べ方の総数が ${}_{10}C_2 \times {}_8C_3$ であることに注意する. 従って任意の k, l に対し, 求める確率 $p(k, l)$ は

$$p(k, l) = {}_{10}C_k q^k {}_{10-k}C_l p^l (1-p-q)^{10-k-l}$$

となる.

- (5) $p(k, l)$ の定義から, この確率を可能な全ての k, l について和をとった p_{total} の値は 1 になることがすぐにわかるが, ここでは実際にこの和を計算することにより, これを確かめておくことにする. ヒントに与えた 2 項定理を用いると

$$\begin{aligned} p_{\text{total}} &= \sum_{k=0}^{10} \sum_{l=0}^{10-k} {}_{10}C_k q^k {}_{10-k}C_l p^l (1-p-q)^{10-k-l} \\ &= \sum_{k=0}^{10} {}_{10}C_k q^k \sum_{l=0}^{10-k} {}_{10-k}C_l p^l (1-p-q)^{10-k-l} \\ &= \sum_{k=0}^{10} {}_{10}C_k q^k \{p + 1 - p - q\}^{10-k} \\ &= \sum_{k=0}^{10} {}_{10}C_k q^k (1-q)^{10-k} \\ &= \{q + 1 - q\}^{10} = 1^{10} = 1 \end{aligned}$$

となる.

演習問題 1

その状態遷移図が図 4 で表現される通信路について以下の問いに答えよ.

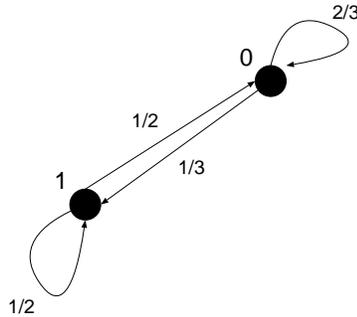


図 4: ここで考える状態遷移図.

- (1) 状態遷移行列: P を求めよ.
- (2) この通信路に 1 を入力したとき, 出力が 0, 1 である確率: P_0, P_1 をそれぞれ求めよ.
- (3) P^2 でその遷移確率が定義される通信路を図 4 にならって図示せよ.

- (注 1) このようにして毎回出題されるレポート提出の積み重ねにより, 最終成績の約 40% が決まります.
- (注 2) 締め切りは次回の講義開始前. レポート用紙には氏名, 学籍番号を必ず書くこと.
- (注 3) レポートは次回に返却できない場合があります. 自分の作成したレポートをそれ以後 (レポートが返却されるまで) の自習に使う者は, 必ずレポートのコピーをとってから提出してください.