北海道大学
Hokkaido University

第2回 情報量とエントロピー（1）

工学部 情報エレクトロニクス学科

情報科学研究院 情報理工学部門
大規模知識処理研究室

堀山 貴史

教科書の構成

- 中村、喜田、湊、廣瀬 著
  基礎から学ぶ 情報理論（第2版）、ムイスリ出版

今回
＋
次回

1. 情報理論とは
2. **情報量**と**エントロピー**

概要と準備

3. 情報源のモデル
4. 情報源符号化の限界
5. 情報源符号化法

情報源とその符号化

6. 通信路のモデル
7. 通信路符号化の限界
8. 通信路符号化法

通信路とその符号化

文献の情報

- C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October, 1948.
- 『通信の数学的理論』シャノン，ウィーバー（ちくま学芸文庫）

- Cover & Thomas (2006) 教科書の文献(9)
- 『情報理論 –基礎と広がり–』Cover & Thomas（共立出版）

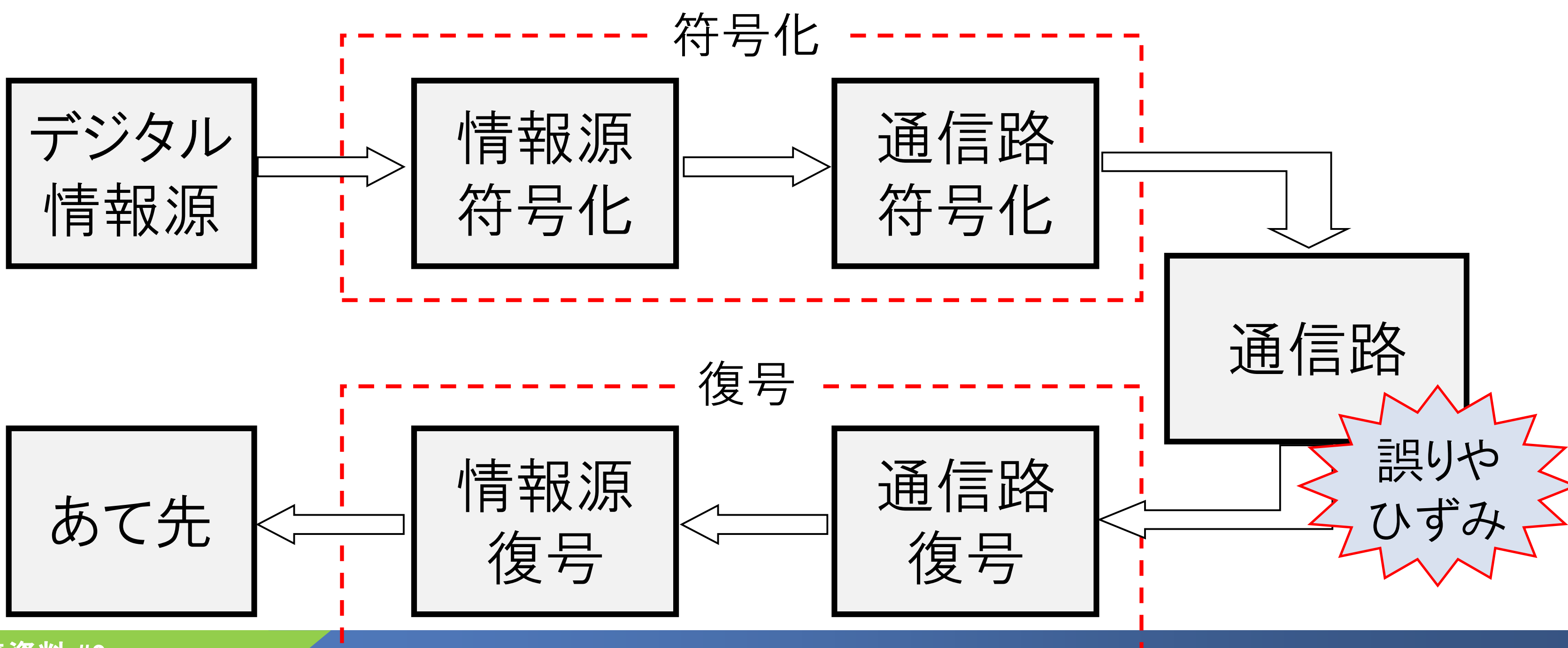
前回の復習: 情報理論が取り組む4つの課題

【課題1】

【課題2】

【課題3】

【課題4】



前回の復習： 情報

- 記号の系列（データ）を受け取ることで、
  受け手の確率分布（知識）が変化する、その変化

- 札幌で暮らしている人
- 冬の朝に起きて、今日の天気が**雪**と知っても、驚かない
  - –いつも雪だから （受け手の知識は、あまり変わらない）
- **雨**が降ったら、アレって思う
  - –レアな事象だから （この情報で、受け手の知識は、大きく変わる）

地球の気象
（情報の発信源）

今日の天気

晴、曇、雨、雪

受け手

天気の統計
（統計的知識）

晴	5.5 %
曇	1.2 %
雨	0.2 %
雪	93.1 %

前回の復習： 情報（レポート課題）

- 記号の系列（データ）を受け取ることで、
  受け手の確率分布（知識）が変化する、その変化

情報の量
＝知識の変化量

- 外の天気は知らず、窓を開けると、外の天気が**雪**だった　**小**
- 外の天気は知らず、窓を開けると、外の天気が**雨**だった　**大**
- **もともと外の天気を知っていて**、窓を開けると外の天気が雨だった

窓を開けても、**知識が変化しない**（もともと、雨 100 %、晴 曇 雪 0 %）
つまり、窓を開けることで得られた情報量は、**0**

地球の気象
（情報の発信源）

今日の天気

晴、曇、雨、雪

受け手

天気の統計
（統計的知識）

晴	5.5 %
曇	1.2 %
雨	0.2 %
雪	93.1 %

一つの結果を知ったときの情報量

確率 p の事象の生起を知ったときに得られる**情報量**を $I(p)$ とする
$I(p)$ は、次のような性質を満たすべき

① $I(p)$ は $0 < p \leq 1$ で単調減少な関数である

② 確率 p_1, p_2 で起こる二つの互いに独立な事象が同時に起こる確率 $p_1 p_2$ について $I(p_1 p_2) = I(p_1) + I(p_2)$

③ $I(p)$ は $0 < p \leq 1$ で連続な関数である

今日の天気は晴れ、コインを1回投げたら表

情報量の加法性

これらを満たす関数 $I(p)$ は

$$I(p) = \boxed{}$$

という形しかありえない（ただし $a > 1$）

証明は省略
（詳しくは教科書で）

対数の定義

定義A.1:

$a \neq 1$ を正の実数とする．このとき，任意の実数 $x > 0$に対し

$$x = a^p$$

を満たす実数 p が唯一に決まる．これを $p = \log_a x$ と書き，a を底とする x の **対数** (logarithm) と呼ぶ．

つまり、a を何乗したら x になるか

定義A.2:

$$e = \lim_{t \to 0}(1+t)^{\frac{1}{t}} = 2.718281828459 04 \cdots$$

で定義される定数e をネイピア数（またはオイラー数）という．

- 底が10のときを常用対数 (Log x)
- 底がe のときを**自然対数** (ln x)
- 底が2のときを**二進対数** (lg x， log x)

情報理論では二進対数が良く用いられる

対数の基本的な定理

定理A.1：$a \neq 1$ を正の実数とする．任意の正の実数x, y と任意の実数b に対し，以下が成り立つ．

(1) $a^{\log_a x} = x$

(2) $\log_a a = 1$

(3) $\log_a 1 = 0$

掛け算が足し算に！

割り算が引き算に！

(4) $\boldsymbol{\log_a xy = \log_a x + \log_a y}$ (4’) $\boldsymbol{\log_a \frac{x}{y} = \log_a x - \log_a y}$

(5) $\boldsymbol{\log_a x^b = b \log_a x}$ (5’) $\log_a \frac{1}{x} = \log_a x^{-1} = -\log_a x$

(6) $a > 1$ かつ$x < y$ のとき，$\log_a x < \log_a y$

(7) $a > 1$ のとき，$\lim_{x \to +\infty} \log_a x = +\infty$

定理A.2：正の実数$a, b\ (\neq 1)$ および任意の正の実数x に対し，

$$\log_a x = \frac{\log_b x}{\log_b a}$$

底の変換！

が成り立つ．

情報量 （さっきの話の続き）

定義2.1

確率 p で生起する事象が起きたことを知ったときに得られる情報量 $I(p)$ を**自己情報量**と呼び,

と定義する. ただし, a は $a > 1$ の定数とする.

情報量 （さっきの話の続き）

定義2.1

確率 p で生起する事象が起きたことを知ったときに得られる情報量 $I(p)$ を**自己情報量**と呼び,

と定義する. ただし, a は $a > 1$ の定数とする.

$a = 2$ の場合，単位は**ビット**（bit）という

- 自然対数で計るときは**ナット**（nat） 1 nat≒1.443 bit
- 10を底とする対数で計るときは**ハートレー**（Hartley）
- もしくは**ディット**（dit）または**デシット**（decit） 1 Hartley≒3.322 bit

確率1/2で生じる結果を知ったときの情報量 ＝1 [bit]

<u>簡単な例題</u>: サイコロを1回振ったときの出目を知ったときに得られる情報量は何ビットか答えよ．ただし，サイコロの各出目が得られる確率はすべて等しく1/6とする.

平均情報量

M個の互いに排反な事象$a_1, a_2, \cdots, a_M$ が起こる確率を $p_1, p_2, \cdots, p_M$ とする（ただし，$p_1 + p_2 + \cdots + p_M = 1$）．
このうち1つの事象が起こったことを知ったときに得る情報量は $-\log_2 p_i$ であるから，これを平均した期待値$\bar{I}$ は，

$$\bar{I} = p_1(-\log_2 p_1) + p_2(-\log_2 p_2) + \cdots + p_M(-\log_2 p_M)$$

$$= \boxed{}$$

となる．これを**平均情報量**（単位はビット）という．

エントロピー

定義2.3

確率変数 X がとりうる値が$x_1, x_2, \cdots, x_M$ とし，X がそれぞれの値をとる確率が$p_1, p_2, \ldots, p_M$（ただし，$p_1 + p_2 + \cdots + p_M = 1$）であるとき，確率変数 X の**エントロピー**を

$$\boldsymbol{H(X)} = -\sum_{i=1}^{M} p_i \log_2 p_i$$

ビットと定義する．

エントロピー

定義2.3

確率変数 X がとりうる値が$x_1, x_2, \cdots, x_M$ とし，X がそれぞれの値をとる確率が$p_1, p_2, \ldots, p_M$（ただし，$p_1 + p_2 + \cdots + p_M = 1$）であるとき，確率変数 X の**エントロピー**を

$$\boldsymbol{H(X)} = -\sum_{i=1}^{M} p_i \log_2 p_i$$

ビットと定義する.

<u>例題2.1</u>：　偏りのないコインを2回投げて表の出た枚数を確率変数 X とする. このとき，X のエントロピー$H(X)$ は何ビットか？

$$H(X) = -\frac{1}{4}\log_2\frac{1}{4} - \frac{1}{2} \times \log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4}$$

$$= 2 \times \frac{2}{4} + \frac{1}{2} = 1.5 \quad \text{(ビット)}$$

X	0	1	2
確率	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Try レポート課題

平均情報量の直感的意味は

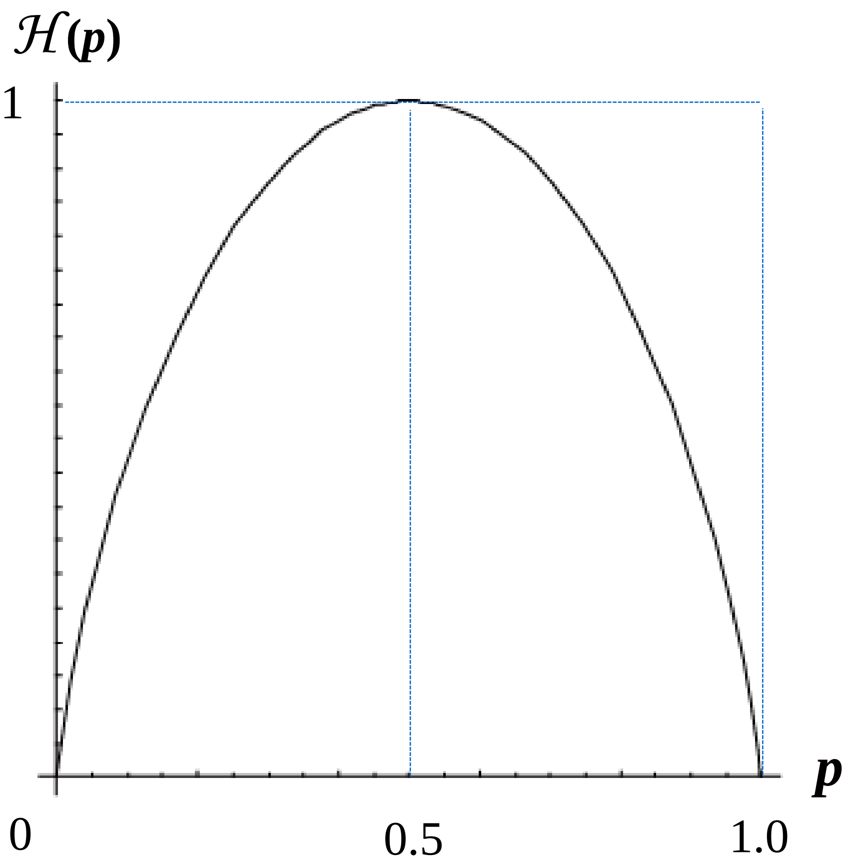
- 北海道では、冬の交通情報は重要度が高い
  - 夏はめったに障害が発生しないが、冬はしばしば発生する
  - 夏は、交通情報を聞き逃したとしても、あまり影響はない
  - 冬は、交通情報を聞かないと、ひどい目に遭うことが多い

  → 冬の方が交通情報の平均情報量が大きい
- 一般に、確率分布の偏りが非常に大きい（片方が 1 に近く、片方が 0 に近い）ときは平均情報量が小さく、確率分布が均一（五分五分）のときに平均情報量が最大になる
- **エントロピーと平均情報量は、完全に同じ形をしている**
  - 元々、エントロピーは熱力学の分野で、確率現象の不確かさ、曖昧さを表す尺度として使われていた
  - 情報を受け取ると曖昧さが減るので、受け取った量だけエントロピーは減って、0 に近づく**（受け取った情報量＝エントロピーの変化）**

M = 2 (確率p, 1-p) のときのエントロピー

$$H(X) = -p \log_2 p - (1-p) \log_2 (1-p)$$
$$= \mathcal{H}(p)$$

エントロピー関数



$\mathcal{H}$ を「エントロピー関数」と呼ぶ。
エントロピー関数は、M = 2 のときに限り
数式を短縮して書くための単なる記法で、
それ以上の特別な意味はない。

2つの事象が五分五分で生起

$p = 0.5$ のとき**最大値** 1.0 (bit)

$p = 0$ または 1.0 のとき**最小値** 0 (bit)

1つの事象が確定して生起

エントロピーの性質

定理2.1

M個の値をとる確率変数Xのエントロピー$H(X)$は次の性質を満たす.

(1) $\boxed{}$

(2) $H(X)$が最小値0となるのは，ある値をとる確率が1で，他の$M-1$個の値をとる確率がすべて0のときに限る．すなわち，**Xのとる値が初めから確定している場合のみ**である.

(3) $H(X)$が最大値$\log_2 M$となるのは，**M個の値がすべて$1/M$で等しい場合**に限る.

例

- 英文字が 1 つ発生する確率変数 X を考える
- すべての文字が等確率で発生するとしてエントロピーを計算すると、

$$H(X) = \log_2 26 = 4.70 \text{ (bit)}$$

- 実際は表のような偏りがあるので、

$$H(X) = -\sum_{i=1}^{26} p_i \log_2 p_i = 4.17 \text{ (bit)}$$

となる

- 英語の文章は1文字ごとに**独立ではなく**、隣接する文字同士に**相関関係**があるため、実際の冗長度はもっと大きい（エントロピーは3.53程度であると**推定されている**）

英文における文字の出現確率

文字	確率	文字	確率	文字	確率
A	8.29%	J	0.21%	S	6.33%
B	1.43	K	0.48	T	9.27
C	3.68	L	3.68	U	2.53
D	4.29	M	3.23	V	1.03
E	12.08	N	7.16	W	1.62
F	2.20	O	7.28	X	0.20
G	1.71	P	2.93	Y	1.57
H	4.54	Q	0.11	Z	0.09
I	7.16	R	6.90		

スペースや句読点は考えないことにする

情報量の意味（まとめ）

- 個々の情報量は、確率が低い事象ほど大きい値となる。
- 平均情報量(=エントロピー)は、確率分布の偏りが大きいほど小さい値となる。
  - 2値の事象では、1/2ずつのときに最大値1(bit)となる
  - 4値の事象では、1/4ずつのときに最大値2(bit)となる
  - M値では、1/Mずつのとき最大値 $\log_2 M$ (bit)となる。
  - 何値であっても、どれか1つの事象が100%で残りが0%のときに最小値 0 となる。
- エントロピーは「答の当てにくさ」を表す数値とも言える。
  - 分布の偏りが大きければ当てやすく、偏りがないほど当てにくい。
  - ヒントが与えられると「当てにくさ」は低下する。
  - 答が完全にわかってしまうと「当てにくさ」は0になる。
  - 平均情報量＝答を全く知らない状態から、エントロピーを0に下げるために必要なヒントの情報量

2つ以上の確率変数を扱う場合

2つの事象が**独立**

例1）

- 確率変数 X: 今日の天気　X = 晴 / X = 雨
- 確率変数 Y: 今日コインを1回投げた結果　Y = 表 / Y = 裏

ここから話がややこしく
なってくるので、あわてず
何度か繰り返して
ゆっくり学習してください

・今日の天気は晴
・コインを1回投げたら表

2つ以上の確率変数を扱う場合

例1）

2つの事象が**独立**

- 確率変数 X: 今日の天気　X = 晴 / X = 雨
- 確率変数 Y: 今日コインを1回投げた結果　Y = 表 / Y = 裏

例2）

2つの事象に**相関関係がある**

- 確率変数 X: 今日の天気　X = 晴 / X = 雨
- 確率変数 Y: 今日のアイスの売り上げ

Y = よく売れた / Y = 売れない

・今日の天気は晴
・今日アイスがよく売れた

例1

二つの確率変数X, Yを考える. Xは$x_1, x_2, \ldots, x_{M_X}$の値をとり, Yは$y_1, y_2, \ldots, y_{M_Y}$の値をとるものとする. 確率変数の組$(X, Y)$の値が$(x, y)$ となる**結合確率分布**を$P(x, y)$ と書く

例1

二つの確率変数X, Yを考える. Xは$x_1, x_2, \dots, x_{M_X}$の値をとり, Yは$y_1, y_2, \dots, y_{M_Y}$の値をとるものとする. 確率変数の組(X, Y)の値が(x, y)となる**結合確率分布**を$P(x, y)$と書く

$P(x,y)$		Y: コイン 表	Y: コイン 裏	$P(x)$
X: 天気	晴	1/3	1/3	2/3
	雨	1/6	1/6	1/3
$P(y)$		1/2	1/2	

組(X, Y)をまとめて考えると, 4つの値をとる確率変数 Z のエントロピー $H(Z)$ として考えることができる

(晴、表)
(晴、裏)
(雨、表)
(雨、裏)

結合エントロピー

定義2.5

確率変数 X と Y の結合エントロピー $H(X,Y)$ は,

$$H(X,Y) = -\sum_{i=1}^{M_X}\sum_{j=1}^{M_Y} P(x_i, y_j)\log_2 P(x_i, y_j)$$

により定義される. これを**結合エントロピー**と呼ぶ.

ただし, $\{x_1, x_2, \dots, x_{M_X}\}$ および $\{y_1, y_2, \dots, y_{M_Y}\}$ は, それぞれ X と Y が取りうる値の集合とする.

例1

- (X,Y) = (晴，表)、(晴，裏)、(晴，表)、(晴，裏) の
  結合エントロピー $H(X,Y)$
  $-\left(\frac{1}{3}\right)\log\frac{1}{3}-\left(\frac{1}{3}\right)\log\frac{1}{3}-\left(\frac{1}{6}\right)\log\frac{1}{6}-\left(\frac{1}{6}\right)\log\frac{1}{6}$ = 約 19.2 (bit)
- $H(X)$ = 約 0.92，$H(Y) = 1.00$ なので、
  $H(X,Y) = H(X) + H(Y)$ が成り立つ

独立事象だから

$P(x,y)$		Y: コイン 表	裏	$P(x)$
X: 天気	晴	1/3	1/3	2/3
	雨	1/6	1/6	1/3
$P(y)$		1/2	1/2	

例2

- $P(x)$, $P(y)$ は、例1と同じ → $H(X)$, $H(Y)$ も例1と同じ
- しかし、(X,Y) の事象の生起確率には、偏りがある
  結合エントロピー $H(X,Y) = 1.46$ (bit)
- つまり、 $H(X,Y) < H(X) + H(Y)$

X と y が**独立**ではない（相関関係がある）から

$P(x,y)$		Y: アイスの売り上げ 売れた	売れない	$P(x)$
X: 天気	晴	1/2	1/6	2/3
	雨	0	1/3	1/3
$P(y)$		1/2	1/2	

結合エントロピーの性質

定理2.2

確率変数 X と Y の結合エントロピー $H(X,Y)$ に対し,

$$0 \leq H(X,Y) \leq H(X) + H(Y)$$

が成り立つ. また $H(X,Y) = H(X) + H(Y)$ となるのは,
X と Y が **独立のときのみ** である.

今日のまとめ

- **自己情報量**
  - 確率 p で生起する事象が起きたことを知ったときに得られる情報量

$$I(p) = -\log_a p$$

- **平均情報量**
  - M個の互いに排反な事象$a_1, a_2, \cdots, a_M$ のうち１つの事象が起こったことを知ったときに得る情報量（期待値)

$$-\sum_{i=1}^{M} p_i \log_2 p_i$$

- **確率変数** X のエントロピー（= 平均情報量）

$$\boldsymbol{H(X)} = -\sum_{i=1}^{M} p_i \log_2 p_i$$

- **確率変数** X, Y の結合エントロピー
  - X と Y を組 (X, Y) にして、この組についてのエントロピー

$$H(X, Y) = -\sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} P(x_i, y_j) \log_2 P(x_i, y_j)$$