講義「情報理論」

第2回　情報量とエントロピー（1）

情報理工学専攻 情報知識ネットワーク研究室
喜田拓也

2017/6/12

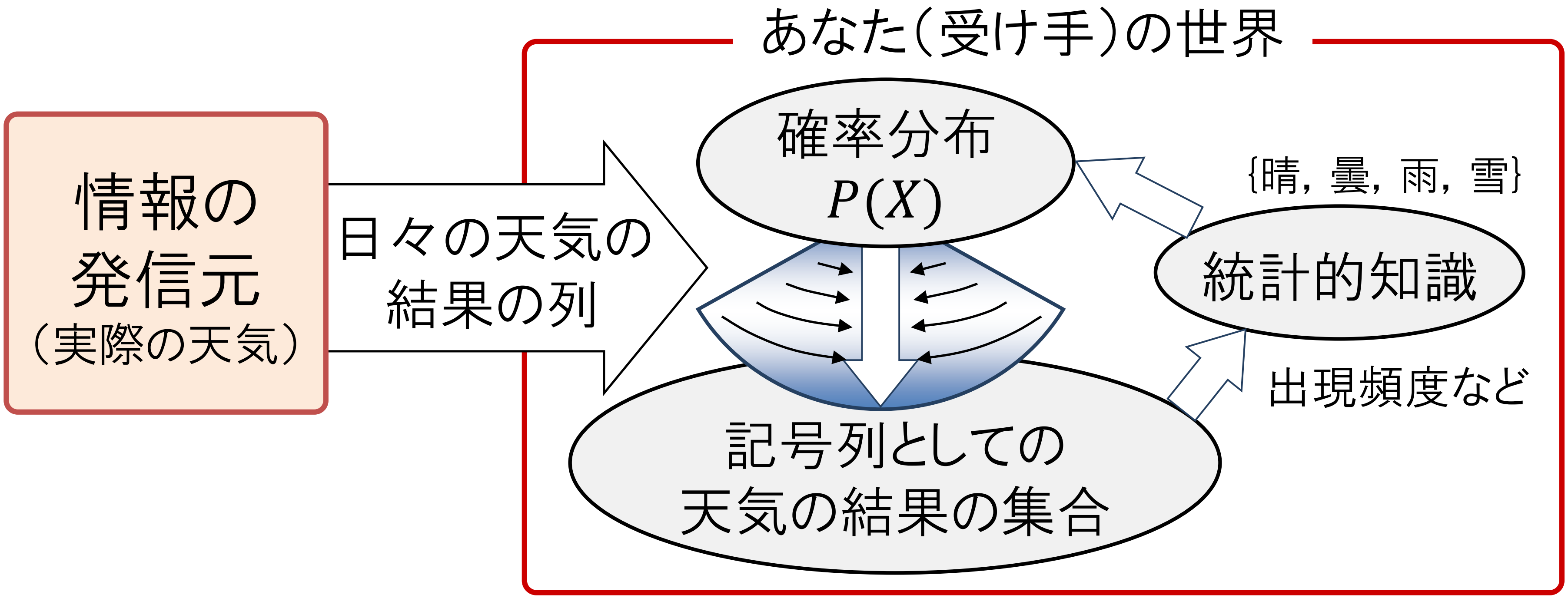
情報理論での情報伝達のモデル（おさらい）

情報を「記号の列」だと割り切って取り扱う

その意味内容には立ち入らない

送信される情報は記号{晴，曇，雨，雪}の列

あなた（受け手）の世界は，記号（列）の集合とその統計的知識



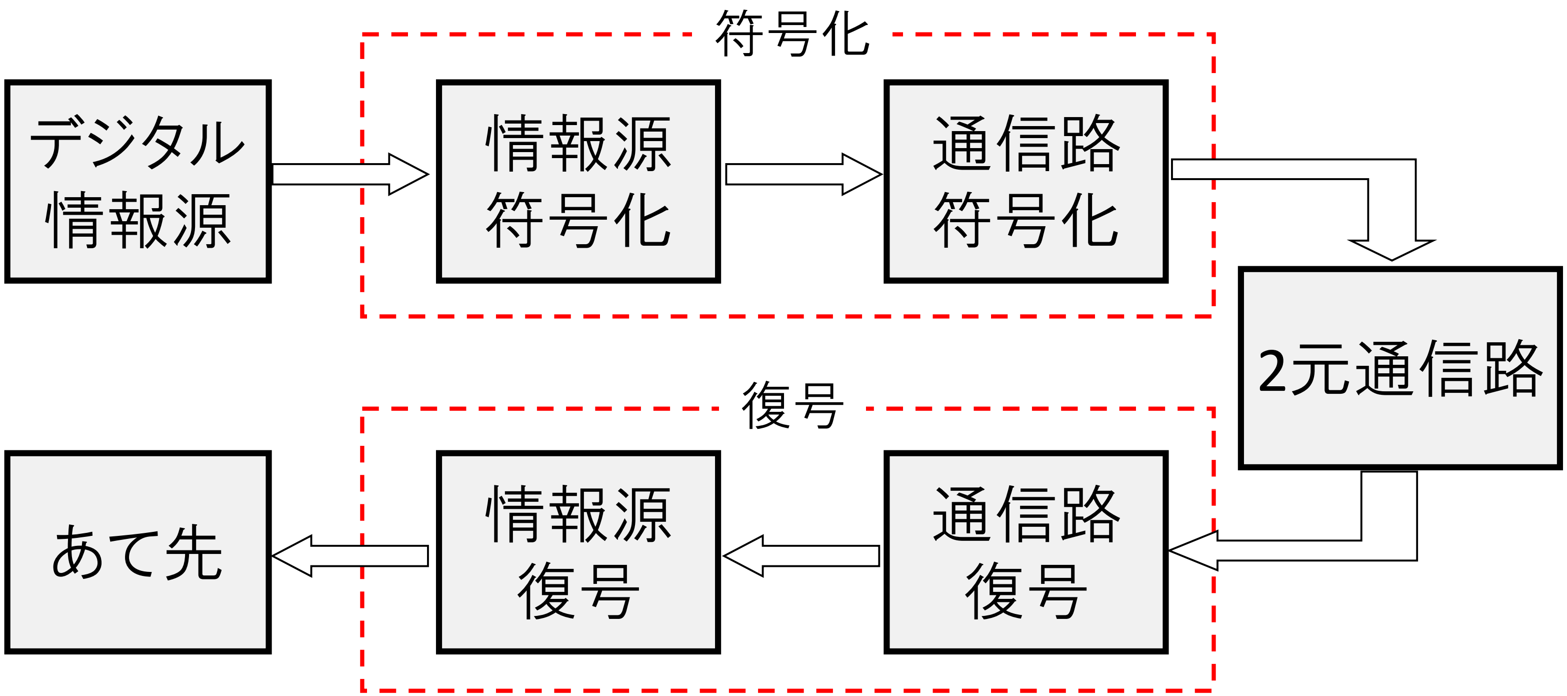
情報理論が取り組む4つの問題(おさらい)

【問題1】できるだけよい情報源符号化法(復号法)を見出すこと

【問題2】情報源符号化の限界を知ること

【問題3】できるだけよい通信路符号化法(復号法)を見出すこと

【問題4】通信路符号化の限界を知ること



今日の内容

確率について基本的なお話

付録A 対数とその性質

2.1 情報量

2.2 エントロピー

確率とは

古典的には

「ある事象についての起こりやすさの指標」
もとは，賭け事の戦略や，金銭の公平な分配を考えるために用いられた概念

「1個のサイコロを振って，奇数だったら私の勝ち，偶数だったらあなたの勝ち．参加費は100円．勝ったら180円をあげるよ．勝負する？」

事象，起こりやすさの指標って？

偶然現象において，起こりうる事柄を事象という

それぞれの事象が起こりうる可能性を，すべての起こりうる事象に対する割合として表すとき，その値を確率という

古典的な確率の定義

$$P(A) = \frac{r}{n}$$

r：事象Aが起こる場合の数

n：すべての場合の数

偶然現象って何？

サイコロ振りだって物理現象でしょ？
ロボットが振ったらいつも同じじゃん

現代の確率の考え方

「サイコロの目がどう出るかは最初のサイコロの持ち方とその振り方によって決まる．そこには何の偶然性もない．ただし，その因果関係はあまりにも複雑なので，結果を予想するのはほぼ不可能である．・・・（中略）．このような予測しにくいこと，制御しにくいことを「偶然である」とみなすのが確率統計の出発点である．」

（金谷健一著，「確率統計を学ぶにあたって」岡山大学工学部講義資料）

現代の確率論は，「確率」が何を意味するかは問わない．「確率」が満たすべき数学的な性質を規定し，その性質から導かれる定理を論じる．

公理主義

確率の公理による定義　by コルモゴルフ(Kolmogorov)

ある空でない集合 Ω（結果の集合，標本集合）について，その任意の部分集合を事象という.

事象全体の集合を$\mathcal{F}(=2^{\Omega})$とする.

写像$P\colon \mathcal{F} \to [0,1]$が任意の$A \in \mathcal{F}$について次を満たすとき，$P(A)$を事象$A$の確率という.

$Pr(A)$と書くことも

1. $0 \leq P(A)$
2. $P(\Omega) = 1$
3. $A_1, A_2, \ldots \in \mathcal{F}$について$A_i \cap A_j = \emptyset\ (i \neq j)$ならば，$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

$\emptyset$:（ヌル）空集合

確率の三公理

公理から導かれる確率の基本的性質

1. $P(\emptyset) = 0$

2. $P(\bar{A}) = 1 - P(A)$

$\bar{A}$は事象Aが起こらない事象（余事象）

3. 事象$A, B \in \mathcal{F}$に対して

$$A \subset B ならば P(A) \leq P(B)$$

4. 任意の事象$A \in \mathcal{F}$に対して$P(A) \leq 1$

$0 \leq P(A) \leq 1$

5. 事象$A, B \in \mathcal{F}$に対して

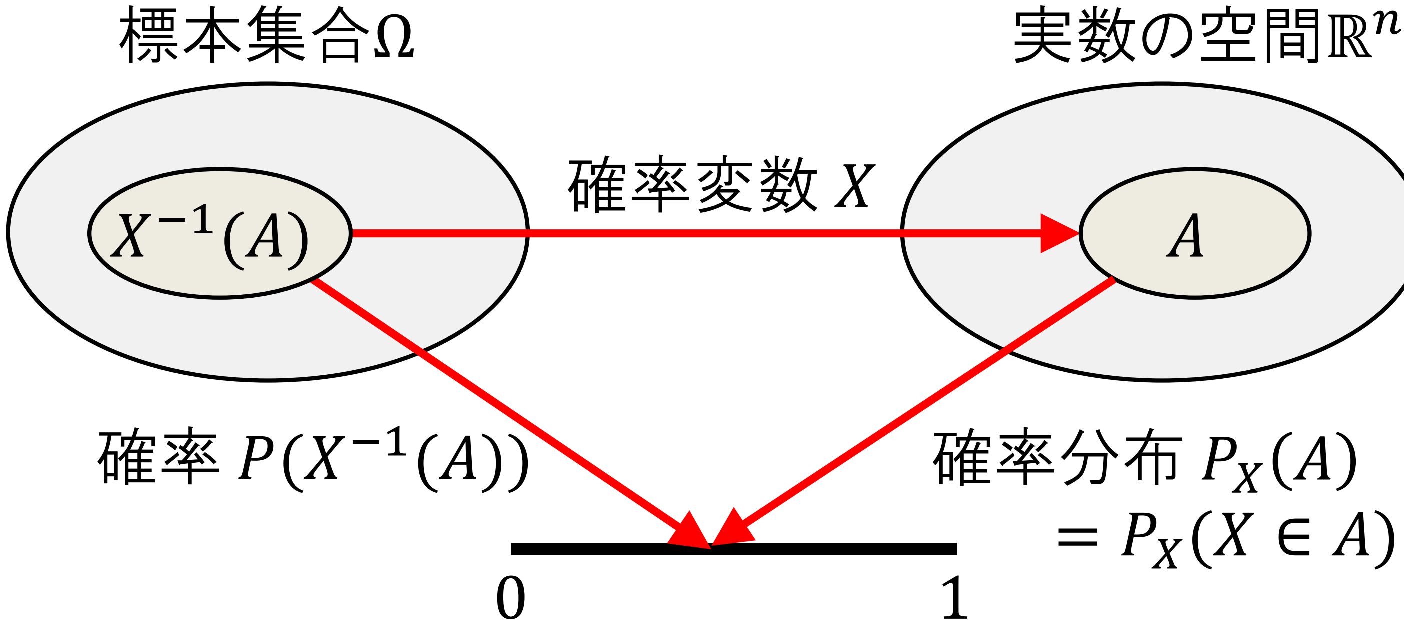
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

確率の加法定理

確率変数と確率分布

確率変数Xとは，標本集合Ωから実数空間$\mathbb{R}$（$\mathbb{R}^n$）への（確率と関連付いた）写像である

確率変数Xがある値になる確率を確率分布という



$$P_X(A) = P(X^{-1}(A))$$

なんでこんなに面倒な議論をするの？

A. 抽象的な概念を抽出して議論することで,
様々なケースに適用できるから

離散的な集合だけでなく, 連続的な集合の上でも
同じようにして確率を定義できる

確率的な性質さえ満たすなら, 何でもいい
→ 主観的な確信の度合いとして確率を用いても
いいじゃないか（ベイズ主義）

直接に統計情報を取れない
事象を推測するのに用いる

まあ難しい話はおいといて

「1個のサイコロを振って，奇数だったら私の勝ち，偶数だったらあなたの勝ち．参加費は100円．勝ったら180円をあげるよ．勝負する？」

各要素が等しく起こりうる場合

標本集合 $\Omega = \{1,2,3,4,5,6\}$
あなたが勝つ事象 $A = \{2,4,6\}$
事象Aが起こる確率 $P(A) = |A|/|\Omega| = \frac{3}{6} = \frac{1}{2}$
得られる賞金額を確率変数Xとすると，その確率分布P_Xは

$$P_X(X = 0) = \frac{1}{2}, \qquad P_X(X = 180) = \frac{1}{2}.$$

得られる賞金額の平均（期待値）$E(X)$は

$$E(X) = \sum_{x \in \{0,180\}} x \cdot P_X(X = x) = 0 \times \frac{1}{2} + 180 \times \frac{1}{2} = 90.$$

Try 問1

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.

定義A.2:

$$e = \lim_{t \to 0}(1 + t)^{\frac{1}{t}} = 2.71828182845904 \cdots$$

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($\mathrm{Log}\, x$),
底がe のときを自然対数($\ln x$),
底が2のときを二進対数($\lg x$)という.

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

(5) $\log_a x^b = b \log_a x$　(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．

ノート A.1

xのaを底とした対数$\log_a x$は，xをa進数表示したときの桁数
(記述長)に対応する．
例えば，ある自然数nの10進数での桁数をmとする．このとき，

$$10^{m-1} \leq n < 10^m \, .$$

各辺の対数をとると

$$m - 1 \leq \log_{10} n < m \, ,$$
$$\therefore \; \log_{10} n < m \leq \log_{10} n + 1 \, .$$

したがって，

$$m = \lfloor \log_{10} n + 1 \rfloor \, .$$

床関数 $\lfloor x \rfloor$:
x 以下の最大の整数

※ Kenneth E. Iverson が1962年に導入

例題: 10進数で 225 は，二進数で何桁か？

$$\log_2 225 = \log_2 5^2 \cdot 3^2 = 2\log_2 5 + 2\log_2 3$$
$$\fallingdotseq 2(2.322 + 1.585) = 7.814$$

11100001

すなわち，$7 \leq \log_2 225 < 8$ なので，8桁である．

$y = \log_a x$ の性質

$a > 1$ のとき，$\log_a x$は単調増加
（x に対してなだらかに増加する）

$x \to +\infty \Rightarrow \log_a x \to +\infty$

$x < y \iff \log_a x < \log_a y$

$0 < x < 1$ の間は負の値をとる

$x \to 0 \Rightarrow \log_a x \to -\infty$

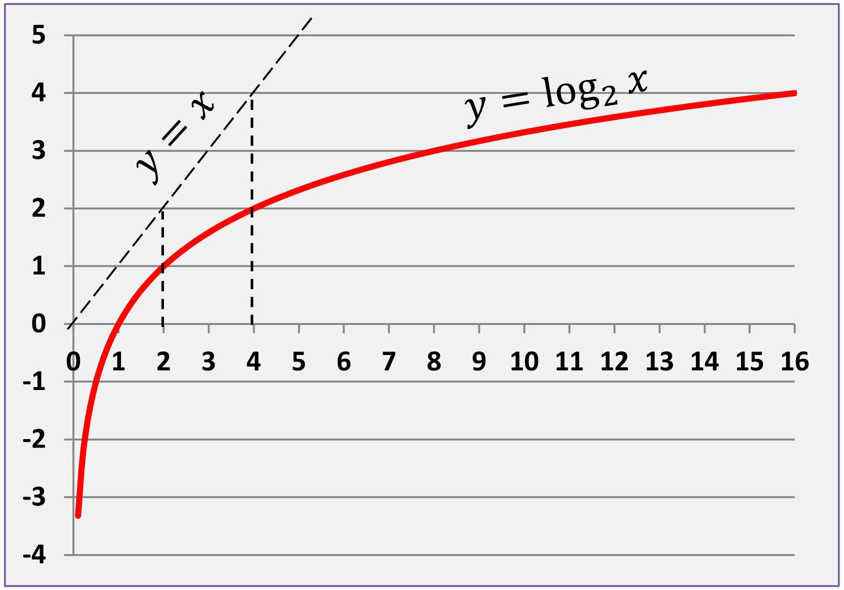


図 $y = \log_2 x$ のグラフ

定理A.3: $\dfrac{d}{dx}\log_e x = \dfrac{1}{x}$.

一般的には，$\dfrac{d}{dx}\log_a x = \dfrac{\log_a e}{x}$.

定理A.4: 正の実数$a \neq 1$ に対して，次が成り立つ.

（1） $y = \log_a x$ の曲線は$y = a^x$ の曲線と直線$y = x$ に関して対称

（2） $\ln x \leq x - 1$

Try 問2，問3

ちょっと休憩

情報には量がある！

確率が高いことを知らされても,
そのニュースは価値が低い

確率1の結果が知らされる → 得られる情報量は 0

情報には量がある！

確率が低いことを知らされたら,
そのニュースは価値が高い

確率が 0 に近い事柄を知らされる → 情報量は大！

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
$I(p)$ とすると $I(p)$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

情報量の加法性

これらを満たす関数 $I(p)$ は

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

という形しかありえない（a は定数）

証明は省略

情報量の定義

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,

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

と定義する. ただし, a $(a \neq 1)$は定数とする.

$a = 2$ 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]

簡単な例題:　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

定義2.2

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，

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

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

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

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

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 練習問題2.1

今日のまとめ

現代における確率の考え方について学んだ

付録A 対数とその性質について復習した

2.1 情報量

確率pで起こる事象の自己情報量 $I(p) = -\log_a p$

2.2 エントロピー

確率変数Xの平均情報量$H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i$

次回テーマ

情報量とエントロピー(2)