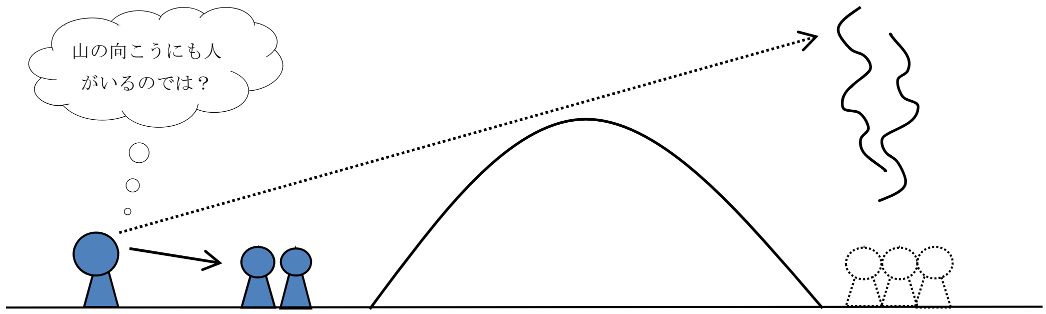
第６章　情報量・エントロピー

１．情報とは

（１）情報と状況

情報とはよく使う言葉です。しかし、「情報とは何ですか？」と聞かれると、返答に困る言葉でもあります。

誰が最初に情報という言葉を使ったでしょうか。情報という言葉は、森鴎外（1862～1922）が講義の中で『戦争論』（著：Karl Von Clausewitz）で使われている言葉「Nachrichten」を“情報”と訳したことが最初だと言われています。



状況　状（意味：かたち、ありさま）…山のこちら側に２人いる。

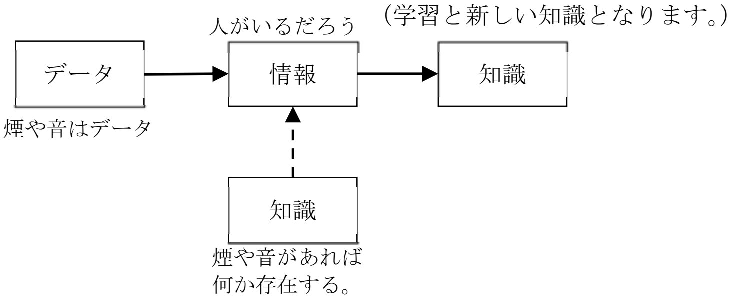
情報　情（意味：心のもと、おもい）…煙がたっているので、たぶん人がいるのではないか。

（例）情報を受け取ると、「不明」なことが「明らか」になります。すなわち、

不明＝あいまいなこと（あいまいさ）。これを表現できれば、情報量として利用できます。

（２）データ・情報・知識の関係

煙のように、音や声はデータです。見た人が「煙があれば何か存在する」との知識があるからこそ、“煙”を情報としてとらえることができます。



２．複雑性、曖昧性と情報量

情報量は、複雑性（さ）、曖昧性（さ）の大小を表すものであるといわれています。私達が複雑さを言うとき、それを調べるのにどのくらいの時間が必要か、または、何回いろいろなものを調べればよいかなどと表現できます。

それでは、10 円が 2 枚あります。それを、机の上に投げたとして、そのうち 2 枚の 10 円が表裏どのようになったか 1 つ 1 つ調べたとします。

「何回調べる」でしょう。

10 円	10 円
表か裏か	表か裏か

1	1
0	0
1 回	2 回

2 回調べます。

それでは、3 枚の 10 円玉ではどうでしょうか。

10 円	10 円	10 円
表か裏か	表か裏か	表か裏か

1	1	1
0	0	0
1 回	2 回	3 回

3 回調べます。

それでは、どちらが表裏でどのような状態になっているか調べるのに、どちらが複雑と言えるでしょうか。当然、3 枚の 10 円玉は 3 回調べることになり、2 枚の 2 回より複雑になります。

そこで、昔学んだ対数を使って考えてみます。

たとえば、2^3＝8 は、$\log_2 8 = 3$ となります。　$(2 \times 2 \times 2 = 8)$

それでは、$\log_2 \frac{1}{8} = \log_2 \frac{1}{2^3} = \log_2 2^{-3} = -3$　(注)$\frac{1}{8} = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2^{-3}} = 2^{-3}$

そこで何回と言うとき、－3 回では、おかしいので、－をつけ、

$$-\log_2 \frac{1}{8} = -\left(\log_2 \frac{1}{2^{-3}}\right) = -(-3) = 3$$

3 回と表すことができます。

処理の予習

対数は、次のように学習しました。（高校）

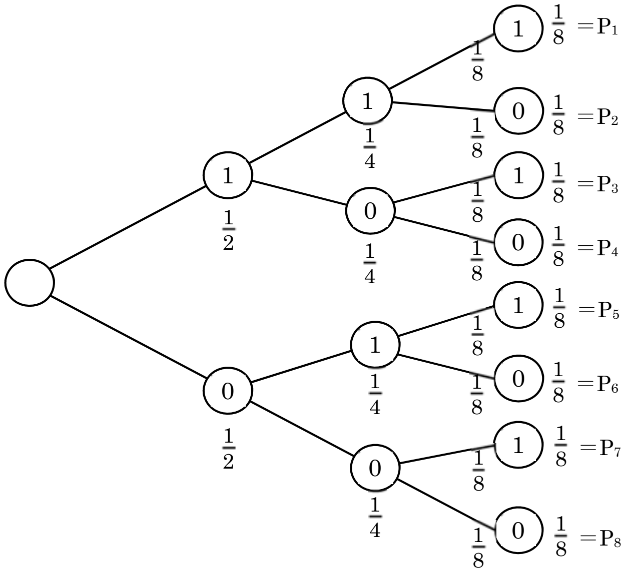
(注)2^3＝8 を $\log_2 8$＝3 と表しました。

・・・8 は 2 の何乗かを示す方法として、$\log_2 8$ のように表す。

（e^m＝xとすると $\log_e x$＝m 、10^n＝yとすると $\log_{10} y$＝n

と表し、e、10 は低と言いましたね。また、e^m の m は指数です。

それでは、3 枚の 10 円玉の表・裏になる確率を考えて下さい。



回	操作の平均	
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$-P_1\log_2 P_1$
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$-P_2\log_2 P_2$
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$-P_3\log_2 P_3$
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$-P_4\log_2 P_4$
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$-P_5\log_2 P_5$
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$-P_6\log_2 P_6$
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$-P_7\log_2 P_7$
$-\log_2\frac{1}{8}$	$-\frac{1}{8}\log_2\frac{1}{8}$	$\underline{\underline{-P_8\log_2 P_8}}$

情報量：全体の平均を出すために各情報 $\log_2 P_i$の量に各確率をかけます。

それでは、これを数式で考えてみてください。

$\log_2\frac{1}{8}=\log_2\frac{1}{2^3}=\log_2 2^{-3}=-3$

そこで、回数として 3 にするために－をかけます。

$-\log_2\frac{1}{8}=3$　になります。

3 枚の 10 円玉を机の上に投げて、平均して、何回で見いだせるかを調べると次のようになります。

平均の回数は$-(\frac{1}{8}\log_2\frac{1}{8}+\frac{1}{8}\log_2\frac{1}{8}+\frac{1}{8}\log_2\frac{1}{8}\cdots\cdots+\frac{1}{8}\log_2\frac{1}{8})=3$

平均して 3 回調べれば、表・裏の状態が明らかになります。

すなわち、3 回調べれば、3 枚の 10 円玉の表・裏の状態が分かります。

そこで、10 円玉が 2 枚、3 枚、4 枚・・・と増えていけば、2 回、3 回、4 回・・・と調べ、どのような状態かの情報を知るのに操作回数、すなわち、あまり性が高くなります。

これを情報量 H と表します。

10円玉は、全体が P＝$\frac{1}{8}$ですが、これが P_1、P_2・・・P_8 に変えるとどうなるでしょう。

図で示すように、$\frac{1}{8}$ の代わりにそれぞれ P_1〜P_8 におきかえます。

$$\mathrm{H} = -\left(\mathrm{P}_1 \log_2 \mathrm{P}_1 + \mathrm{P}_2 \log_2 \mathrm{P}_2 + \cdots\cdots + \mathrm{P}_8 \log_2 \mathrm{P}_8\right)$$

$$\mathrm{H} = -\left(\mathrm{P}_1 \log_2 \mathrm{P}_1 + \mathrm{P}_2 \log_2 \mathrm{P}_2 + \cdots\cdots + \mathrm{P}_n \log_2 \mathrm{P}_n\right)$$

$$\mathrm{H} = -\sum_{i=1}^{n} \mathrm{P}_i \log_2 \mathrm{P}_i$$

情報量 H とは？
P が 1〜n のときは、左の式のようになり、この困難さ・曖昧さを表す量を“情報量 H”と言います。

３．エントロピー

情報量はカテゴリーの数によって最大値が変わってしまいます。

たとえば、2 枚の 10 円玉だと回数はいくつの場合（分類、カテゴリー）があるでしょうか。

$$(\mathrm{H}) = -\left(\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{4}\log_2\frac{1}{4}\right)=2$$

10 円　10 円

もし、表裏のでる確率が 1/2 であれば、表裏を 2 回操作すれば並び方が明らかになります。

$$(\mathrm{H}) = -(\mathrm{P}_1\log_2\mathrm{P}_1+\mathrm{P}_2\log_2\mathrm{P}_2+\mathrm{P}_3\log_2\mathrm{P}_3+\mathrm{P}_4\log_2\mathrm{P}_4) = -\sum_{i=1}^{4}\mathrm{P}_1\log_2\mathrm{P}_i$$

一方、カテゴリー数の違ったものとの比較は困難です。そこで、カテゴリー（j）が取る最大の H を最大 1 になるように変換して利用しています。

カテゴリーjの最大値 H は、次のようになります。（10 円玉 3 つの場合を考えて下さい。）

$$\mathrm{H} = -\log_2\left(-\frac{1}{\mathrm{j}}\log_2\frac{1}{\mathrm{j}}+\cdot\cdot\cdot\cdot\cdot\cdot\cdot\cdot-\frac{1}{\mathrm{j}}\log_2\frac{1}{\mathrm{j}}\right)=\frac{1}{\mathrm{j}}\log_2\frac{1}{\mathrm{j}}$$

たとえば、前の　$-\left(\frac{1}{8}\mathrm{P}_1\frac{1}{8}-\frac{1}{8}\mathrm{P}_2\frac{1}{8}\cdot\cdot\cdot\frac{1}{8}\mathrm{P}_8\frac{1}{8}\right)=-\left(-\frac{3}{8}-\frac{3}{8}\cdot\cdot\cdot-\frac{3}{8}\right)=-\left(-\frac{3}{8}\times 8\right)=3$

これで、10 円玉が 3 枚の場合は、情報量の最大値は 3 になります。

カテゴリー数の違うものと比較をするために、この情報量の最大値で、比較したい対象の情報量を割ります。

これをここでは“規格化する”といいます。

10 円　10 円　10 円

カテゴリー数 3

（情報量の最大値 3）

10 円玉 3 枚（カテゴリー数 3）のとり得る情報量の最大値を 1 としたときの規格化された値を求めるには、情報量を 3 で割ればよいですね。

ここで、少し考えてみて下さい。

3 枚の 10 円玉のうち、もし 1 枚は必ず表が出ることが分かっているとします。すると、操作する回数は、2 回になります（1 枚は必ず表と分かっているので）。この場合の情報量 H は、いくつになりますか。この場合の情報量 H は 2 です。

10 円　10 円　10 円　←1 枚はいつも表

H／最大値＝2／3＝0.666……　　になります。

これを HM とすると、　　HM=0.666

エントロピー

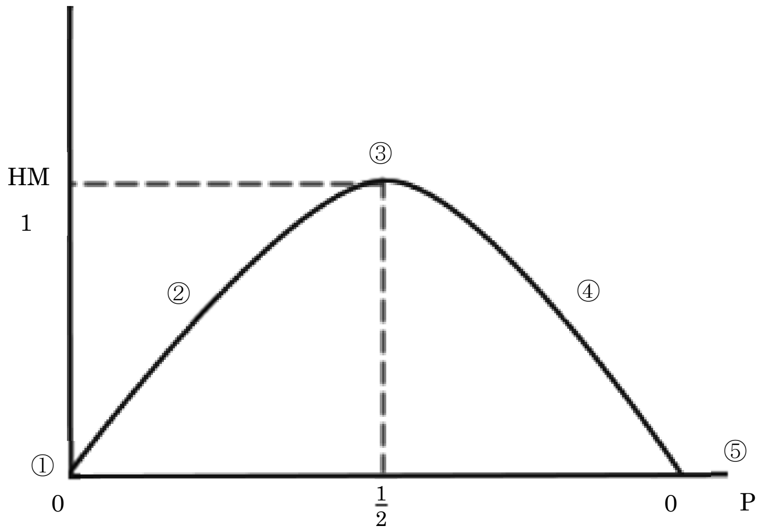
相対情報量

次の HM をエントロピーと言います。

$$HM = \frac{-\sum_{i=1}^{l} \mathrm{P}_i \log_2 \mathrm{P}_i}{-\sum_{i=1}^{l} \frac{1}{l} \log_2 \frac{1}{l}} = \frac{-\sum_{i=1}^{l} \mathrm{P}_i \log_2 \mathrm{P}_i}{\log_2 l}$$

l = カテゴリー数

これをグラフにすると、



これは、教育ではどのような意味をもつでしょうか。

少し考えてみましょう。

①何も知らないときは、エントロピーは 0 ですね。

②先生の話しを聞いて、わからないことが増えてくるとエントロピーは大きくなります。

③知識が増えて、課題（わからないこと・知りたいこと）が最大となります。

④先生の話しを聞く、調べる、教科書を読むなどして、課題が少しずつ明らかになります。

(エントロピーはだんだん小さくなります)

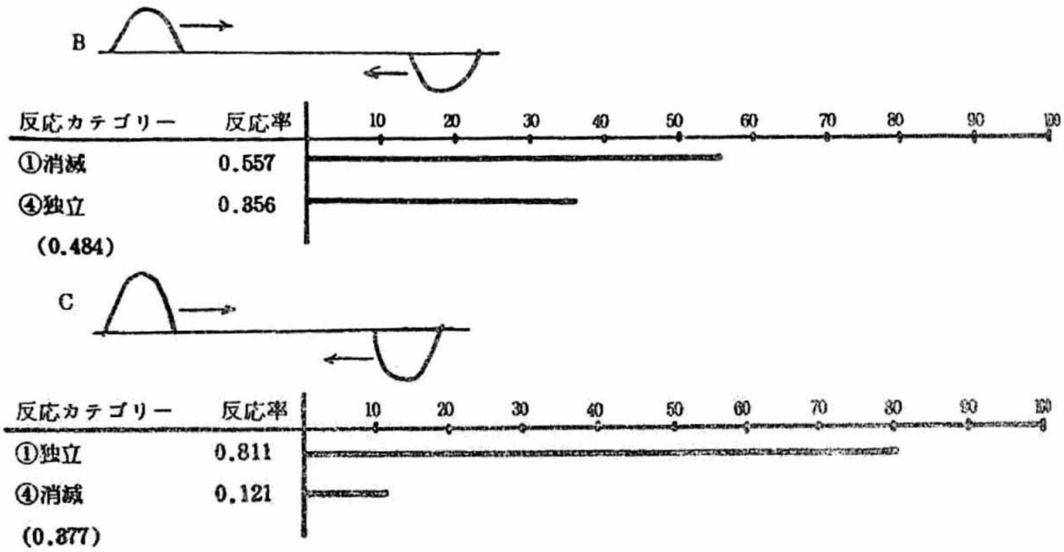
⑤課題がすべて解決（理解）すれば、エントロピーは 0 になります。

普段の授業に当てはめて、考えてみて下さい。

４．カテゴリー分布のエントロピー

問題の誤答や調査（名義尺度）などの処理によくカテゴリー分布が用いられる。このとき、分布の広がり、すなわち答えが複数になってあいまいな状況か、または集中して安定しているかエントロピーを用いて調べられています。

次に示すように、多くの問題のエントロピーの比較がされています。



５．クロス処理したエントロピーの利用

二つの問題の相互の関係として、どのような不安定性（さ）があるか、調べたいことがあります。このときは、二重クロスのエントロピーで計算します。

$$\mathrm{HM} = -\sum_{i=0}^{1}\sum_{j=0}^{1}\frac{\mathrm{P}(\mathrm{A}i、Bj)\log_2\mathrm{P}(\mathrm{A}i、Bj)}{\log_2 4}$$

$i=0$、1　　$j=0$、1　　　　　　$*\log_2 4 = \log_2 2^2 = 2$

問題　ABの関係（事前）、CDの関係（事後）のそれぞれについてエントロピーを求めて下さい。

B(j)(授業前)

A(i)		0	1
	0	10	20
	1	20	50

N=100

P(0.0)=
P(0.1)=
P(1.0)=
P(1.1)=

D(j)(授業後)

C(i)		0	1
	0	0	10
	1	5	85

N=100

P(0.0)=0
P(0.1)=10/100=0.1
P(0.1)=5/100=0.05
P(1.1)=85/100=0.85

2つの問題のクロスの授業前・後のそれぞれのエントロピーを求めて下さい。

□　授業後のエントロピーは

$$\mathrm{HM} = -\{\mathrm{P}(0.0)\log_2\mathrm{P}(0.0)+\mathrm{P}(0.1)\log_2\mathrm{P}(0.1)+\mathrm{P}(1.0)\log_2\mathrm{P}(1.0)+\mathrm{P}(1.1)\log_2\mathrm{P}(1.1)\}\ /2$$

$$= -\{0+0.1\log_2 0.1+0.05\log_2 0.05+0.85\log_2 0.85\}\ /2$$

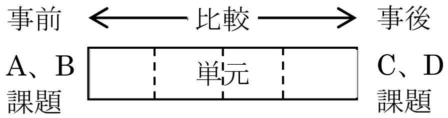
$$= \boxed{\qquad}$$

□　授業前のエントロピー

$\mathrm{HM} = -$

問題　単元の前と後での事前、事後のテストの正誤（1、0）について、相互の関係から、同時確率、条件確率、エントロピー等から何が言えるか考えて下さい。(課題の AB（事前）および、CD（事後）を用いよ）

前頁のクロス処理のエントロピーデータを用いて考えて下さい。



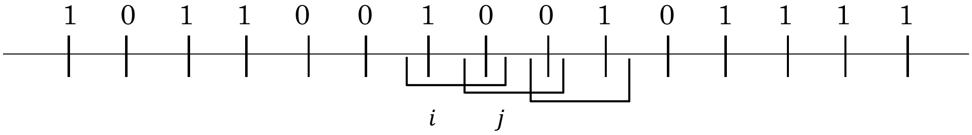
（１）エントロピー（授業前、授業後）からどのようなことが言えますか。

（２）　同時確率　$P(A、B)$、　$P(\bar{A}、B)$の二つを求めて下さい。　$(C、D)$も同様に。

（３）　条件確率　$P(B|A)$、　$P(A|B)$の二つを求めて下さい。　$(C、D)$も同様に。

問題　学習の確かな定着を調べる一つの処理として、この二重クロスのエントロピーが使えます。

児童が一つの事項を繰り返し学習し、最後には「いつもできる」ようになるプロセスを見る一つの手法としても利用できます。



$P(i、j)$でエントロピーを計算し、安定性を検討します。

（i、j）を次々に移動させて計算します。最も使われるのは$P(i、j、k)$のエントロピーです。

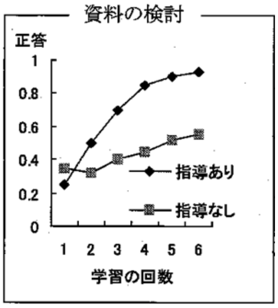
(注)　$P(i、j、k)$は $\mathrm{HM} = \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} P(i、j、k)\log_2 P(i、j、k) / 3$　（$\log_2 \frac{1}{8} = 3$）

○ × × ○ × ○ × × ○ ○ × ○ ○ ○ × ○ ○ × ○ ○ ○・・・・

i　j　k

６．学習指導での利用例

（１）言葉の学習指導への適用例

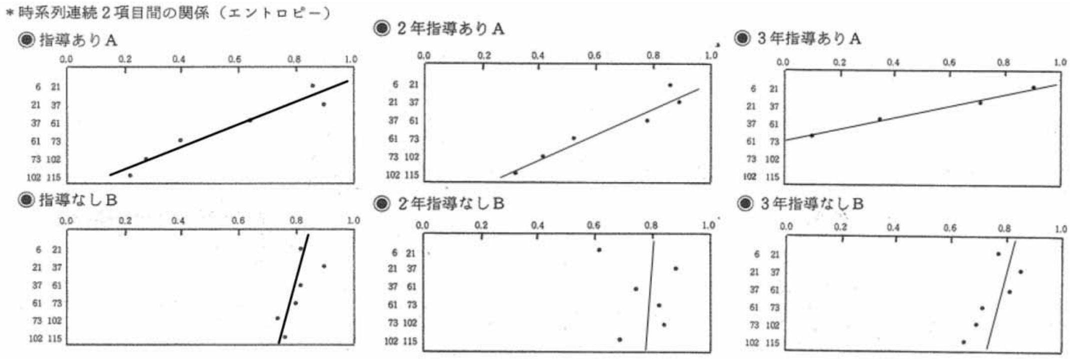


朝の会で毎日の学習プリントを繰り返しした時の正答率の変化を次のグラフに示します。

「指導あり」では、解答後に先生が簡単な説明を行いました。

正答率が上がるものと学習プリントの答え合わせをしたとき簡単な説明をした場合Ａ（上）と答え合わせで終わった場合Ｂ（下）のように違いがあります。

それでは、指導した場合としない場合では、エントロピーにどのような違いがあるでしょうか。



「エントロピーは小さくなると安定です。」

このエントロピーのグラフから何が言えるのでしょうか。

①説明した場合Ａ（上）は

・正答率が高くなる

・学習が安定（エントロピーが小さくなる）

②説明しなかった場合Ｂ（下）は

・正答率が低い

・学習が不安定（エントロピーのが大きく、安定しない）

と言えます。

このことから、教師は学習者の理解が不安定な状態から、安定してできるようにするための学習指導をする必要があります。すなわち、簡単な説明が児童の不安定な学習状態（誤ったけれども、どのように言葉を使ってよいか不安定な状態で「学習プリント」を繰り返し学習するのではなく、このあいまいさをなくす教師の説明が求められます。

説明しないと、正答率は上がりません！！

７．教育でのエントロピーの利用のまとめ

エントロピーは、教育のいろいろな分野で利用されています。事例を調べてみて下さい。教育研究・実践では、エントロピーなどの式を覚える必要はありません。ただし、式を使って考えた「エントロピー、情報の考え方」については、覚えておいてください。

○　情報は、複雑さ・曖昧さを示す。 曖昧さが大きければ、それがわかったとき、その情報の量は大きい。 ∴　分かったことが大きい ○　エントロピーは大きい（１に近い）ときは不安定で、小さいときは安定である。

問題　教育で、どのような使い方ができるでしょうか。

①　授業計画では

②　学習指導では

③　学習結果では

④　授業分析では