

連載 情報理論 情報量について

k06 庄子

○「情報量って何だろう？」

ここでは、簡単に「情報量」というものについて解説していきたいと思っています。

じゃあ、まず「情報」ってそもそも何なのでしょう。

人によって色々と思いつくことがあると思いますが、とりあえず、

- ・何かがそこにあることの知らせ
- ・何かが発生したことの知らせ

を、情報として捉えましょう。



例) “林檎が皿の上にある”というのも、ひとつの情報である。

○情報を数学的に扱うには…？

さて、この“情報”という奴を、数学を使って扱いましょう。数学を使えば、情報に関して様々な性質を導き出すことが出来ます。

しかし、情報はそのままの形では少々扱いづらいようです。上の例だと、「林檎が皿の上にある」という文章の形では、数学で扱うのがちょっと難しいことが容易に想像できるでしょう。

そこで、**「情報」を誰が見ても同じように(つまり客観的に)扱えるようにするため、「情報量」というものを作りましょう。**

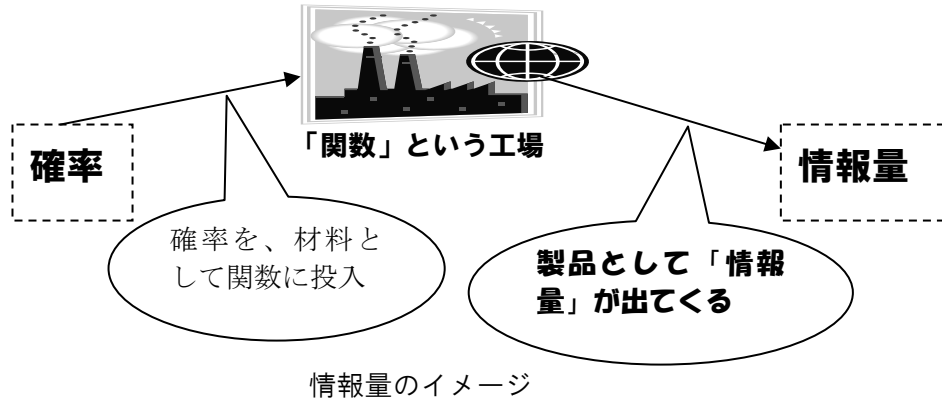
(1) 情報量は確率の関数である

まず、情報量は「誰が見ても同じように」扱えることが大切です。そこで、“情報の性質の中で”誰が見ても同じように扱える部分を考えてみます。

情報の「価値」というのは人によって違うから、誰が見ても同じとはいえないさそうです。では情報の「種類」ではどうかと言うと、これも種類が沢山ありすぎてとても扱えそうなものではなさそうです。

さて、色々と考えてみると面白いものですが…、「その情報の起こる確率」というものはどうでしょうか。“確率”なら、単純に言えば数字ですから、誰が見ても同じように、客観的に扱うことができます。

そこで、情報量を「その物事がある、または発生する確率の関数」として定義しましょう。



(2) 起こりにくい事象の情報量ほど大きくなる。

さて、ここで情報量の大小について考えて見ましょう。

例えば、「東京に雪が降った」という情報と、「東京に謎の煙を出すピンク色のボールが沢山バラ撒かれて、その煙を吸った人が全員引きこもりになった」という情報があったとき、あなたならどっちの情報に「え!？」という衝撃を受けますか？

おそらく、ほとんどの人は「東京に謎の煙を出すピンク色の(略)」の方に「ええっ!？」という衝撃を受けたことと思います。これは、東京に雪が降ったという事象よりも、東京に謎の煙を(略)のほうがはるかに起こりにくいからです。

これを一般化すると、**「起こりにくい情報になるほど情報量が大きくなる」**

きくなる」という事がいえます。

さて、これを少しだけ数学チックに表現してみましょう。

今、二つの事象 a, b があるとします。これらの情報量を、関数 $i(a)$ 、 $i(b)$ で表します。また、 a, b が発生する確率を $P(a)$ 、 $P(b)$ で表します。この時、次の関係が成り立ちます。

$P(a) < P(b)$ の時、 $i(a) > i(b)$ である。

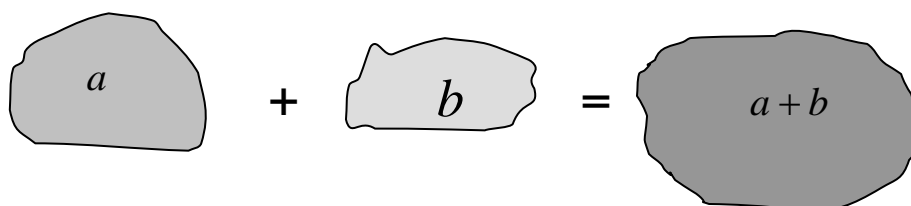
(3) 情報量は加法性を持つ

さて最後に、情報の加法性について考えます。

例えば、あなたが次の2つの情報を同時に受け取った時、その情報量はどれくらいになるでしょうか。

- ・目の前にリンゴがある
- ・後ろにいる人が転んだ

こういう場合、例えば「後ろにいる人が転んだから、目の前にリンゴがある」というようなことはまずありません。つまり、2つの事象の間に何ら関係がないということになります。このように、互いに何ら関係のない2つの事象の情報量を同時に受け取った時、その情報量は2つの情報量の和になります。



さて、これで「情報量が備えるべき条件」が出揃いました。

(1) 情報量は確率の関数である、(2) 起こりにくい事象ほど情報量が大きくなる、(3) 情報量は加法性を持つ、の3つです。

数学の中に、関数と呼ばれるものはたくさんあります。たとえば三角関数 \sin , \cos , \tan 。指数関数 x^n 、などなど。こういった関数の中で、(1)～(3)の条件を満たすものを探してみましょう。特に、(2)と(3)の条件を満たす関数に着目します。

すると、どうやら**対数関数** $\log x$ が該当しそうです。そこで、この対数関数を(1)～(3)の条件に合うよう変形してやります。

(1) 情報量は確率の関数である

事象 a の発生確率を $P(a)$ で表すので、 $\log x$ を $\log P(a)$ とします。

(2) 起こりにくい事象ほど情報量が大きい。

これはつまり、情報量 $i(a)$ は確率の減少関数であるということなので、

式 $i(a) = -\log P(a)$ で表すことができます。

(3) 情報量は加法性を持つ

対数関数はそれ自体で加法性を持つので、特に変更することはありません。

さて、これで遂に情報量の式を求めることができました。

ある事象が存在する確率、または発生する確率を $P(a)$ とする時、その情報量は $i(a) = -\log P(a)$ で表される。

…これで情報量を求めることができましたが、「**で、これがなんの役に立つの？**」という声がビシバシと聞こえてきます。その点についてはまた機会があればこの誌面に掲載したいと思います。それではその時まで、ごきげんよう。