

1.4 情報量の定量的尺度

情報の量的側面がどのように定式化されるか、次のような例を用いて説明しよう。

下記の写真で、今、自分の友人（「山田さん」）がこの路地で先に行ってしまう、左に行ったか右に行ったかわからなくなってしまったとしよう。



単純のために可能性は左に行ったか、右に行ったかのいずれかであるとし、さらに左に行った確率を p 、右に行った確率を $1-p$ とする。すると、このとき、「山田さんはこの路地の奥（左）に行ったよ」ということを教えてもらおうと、その言葉の持つ情報量は、 $\log_2 \frac{1}{p}$ ビットとする。 $p = \frac{1}{2}$ のときだと、その言葉は 1 ビットの情報量であるが、 p がゼロに近づくにつれ、情報量（サプライズ）は増え、逆に 1 に近づくときは、情報量は減る。

自分の友人の行先を聞くことによって得られる平均的な情報量は、

$$p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} \quad \text{ビット}$$

となる。この値は、 $p = \frac{1}{2}$ のとき最も大きい。つまり、左に行った確率も右に行った確率も等しいとき最大になる。

我々は劇場メタファーを設定しているから、そこでは我々の「台本」すなわちモデルに従ってシーンが移り変わっていく。モデルの品質が高ければ、そこで演じられる「劇」は実際に起きることにそれだけ近いことになる。