第四講

この章では、古典情報理論を復習し、それと熱力学の関係をマクスウェルの悪魔のパラドックスを通じて見てみよう。

1　シャノン情報量

この節で述べることは情報科学科の標準的な授業で行われているものを簡単にまとめたものである。

ある確率 p $(0 \leq p \leq 1)$ で、イベント w が起きるとしよう。そのイベントにより得られる情報量を定量化しよう。まず、そのイベントに実際に出くわした時の"びっくり度"を $S(p)$ としよう。宝くじに当たるようなめったに起きないことの場合には"びっくり度"は大きいだろう。言い換えると、びっくり度は p の減少関数だろう。　ここで $S(p)$ に対して次の「加法性」の要請をしよう。確率 p で当選する宝くじと確率 q で当選する宝くじの両方当たったときの"びっくり度"は和、$S(p)+S(q)$ になるとするのである。一方、その両方当たる確率はそれぞれの確率の積 pq になるから、関数方程式

$$S(p) + S(q) = S(pq) \tag{1}$$

が成り立つ。解は定数倍を除いて

$$S(p) = -logp \tag{2}$$

である。

これを一般化しよう。i 番目のイベントが起こる確率を $0 \leq p_i \leq 1$,　$(i = 1, 2, \ldots N)$ としよう。ただし、その総和は規格化されている。$\sum_{i=1}^{N} p_i = 1$。以後、$\{p_1, p_2, \ldots p_N\}$,　$(\sum_{i=1}^{N} p_i = 1)$ を確率分布を呼ぶことにしよう。$\{p_i\}$ と略記することもある。前述したように、加法性から i 番目のイベントが起きたときのびっくり度は $-logp_i$ であるから、"平均的びっくり度"は

$$S(p_1, p_2, \ldots p_N) = -\sum_{i=1}^{N} p_i logp_i \tag{3}$$

である。これをシャノン情報量と呼ぶ。

シャノン情報量の加法性を確認しよう。

$$\begin{gathered} S(\{p_i\}, \{q_i\}) = -\textstyle\sum_{i,j} p_i q_j log(p_i q_j) \\ = -\textstyle\sum_{i,j} p_i q_j log(p_i) - \sum_{i,j} p_i q_j log(q_j) \\ = -\textstyle\sum_i p_i log(p_i) - \sum_j q_j log(q_j) \\ = S(\{p_i\}) + S(\{q_i\}) \end{gathered}$$

例題：　コイン投げ

コインを投げて、確率 p で表（0）確率 $1-p$ で裏（1）としよう。この場合のシャノン情報量 $S(p)$ は

$$S(p) = -plog_2 p - (1-p) log_2 (1-p) \tag{4}$$

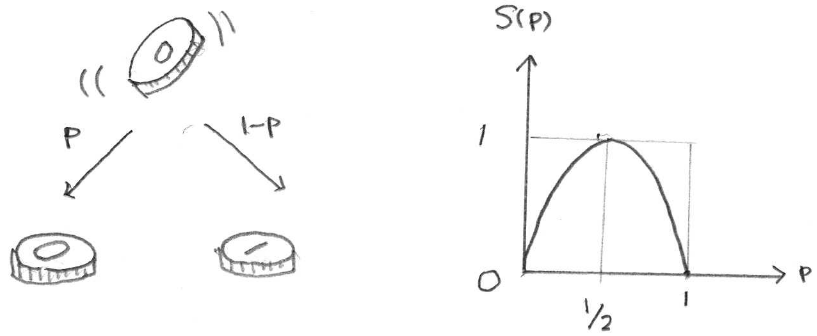

図 1: コイン投げとシャノン情報量

となる。対数の底を 2 に取ったのはあとの便利のためである。特に、$p = 1/2$ でシャノン情報量は最大値 1 を取る。$p = 0$ あるいは $p = 1$ でシャノン情報量は最小値 0 になるが、分かりきった結果には「何の驚きもない」ことと符丁する。別の言い方をすれば、不確定さが大きければシャノン情報量も大きいので、確率現象における結果の不確定さを定量化したものであると言ってもよい。

この節では、シャノン情報量の直観的な導入を行ったが、次節ではシャノンの原論文 [1] に近い形の導入を行う。

1.1 データ圧縮とシャノン情報量

もともと、シャノンはベル電話会社に勤めていて、効率的な通信について研究していた。通信文を 2 進法で送信したとしよう。たとえば、N ビットの文字列、$010010110100\ldots$ などを考えよう。N 個の文字のうち、0 が m 個あるとしよう。そのとき、場合の数は N,m が大きい時、スターリングの公式 $\log(n!) = n\log n - n$ を用いて、

$$W = \begin{pmatrix} N \\ m \end{pmatrix} = \frac{N!}{m!(N-m)!} \approx 2^{NS(p)}. \tag{5}$$

となる。

ここに

$$S(p) = -plog_2 p - (1-p)log_2(1-p) \quad ,p = m/N. \tag{6}$$

このことは、通信文の長さを、N から $NS(p)$, $(\leq N)$ に縮めることができることを意味する。$p = 0$ か $p = 1$ のときには、1 ビットの情報、すなわち 0 あるいは 1 に縮められる。$p = 1/2$ の時には、全く縮められない。$S(p)$ は圧縮度を表す。

これを別の観点から見よう。場合の数が多い程多くの情報を送ることができる。場合に数とは状態の不確定さの度合いでもあるので、情報量の多さと不確定さは裏表の関係にある。