



当選くじ番号の伝え方

A君は、数字当てくじ“ロット7”の当選番号に法則があることを見つけ、今回の当選番号14桁を知ることができた。A君は売り場に行くことができないため、近くに住むBさんに“電報”で伝えることにした。しかし前時代的な電報で送れるメッセージは14文字だけだ。

「ロット クジ カイ！ XXXXXX」この状況で、A君はBさんに当選番号を正確に伝えることができるだろうか？

これは、符号化理論におけるデータ圧縮問題だ。データ圧縮の方法はいろいろある。もっともポピュラーなハフマン法は、一定区間に文字列を区切り、出現確率が高い文字列に短い符号を与えることで、送信文字列を少なくする方法だ。ツイッターや若者同士の会話は、頻度の高い言葉に省略語をあてるので、ハフマン

法の応用と言える。またLZ法は、圧縮するデータ列の中に過去にも出現したデータ列があった時に、過去の出現時点のデータの位置と長さを符号として用いる方法だ。同じ内容のデータ列が頻繁に出現する場合に高い圧縮率が得られる。

しかし、上述の問題の当選番号が“意味のない数字”であった場合は、圧縮法を使ったとしても、問題の答は「ノー」である。元データから圧縮データを作る際の限界を示したのが、有名な「シャノンの

情報圧縮の基本定理¹⁾である。あるデータを0と1の2進数符号で表す際に必要な符号の量をエントロピーと呼ぶ。シャノンの定理は、データのエントロピー量より小さくデータを圧縮して送信することはできないことを示す。もし、それを超えて圧縮した場合には、元に復元することができないのだ。

人間のコミュニケーションについても類似なことが言える。面識のないもの同士は当初、多くの会話が必要だ。その後、コミュニケーションが続いてい

くと、当方と相手の間に想定される“場合の数”（≒エントロピー）が小さくなるので、少ないメッセージで必要なことを伝えることができる。例えば「遅い昼時に、いつもの安い中華屋で待つ」というメールは、当人同士は過去の情報を使い、時間と場所を特定することが可能だ。

ところで、冒頭の問題でBさんは、当選番号は解読できた。が、“ク”を“7”と見誤り、「ロットクジ買い！」を「ロット7次回！」と理解したため、当選金を手に入れそこなった。シャノンの理論には、このようなメッセージの誤送受信を訂正する方法もあるのだが。

（外園 康智）



1) シャノンの理論は、通信路で効率よく情報を伝送するための基礎であり、電話やインターネットなどのインフラを支えている。