1.6 なぜ log か？

生起確率 p の事象が起きたことを知らせる「文」の情報量を $I(p)$ と一般的に表記しよう．これまでの議論では，$I(p)=C\log_e p$ としてきた．なぜそのようにしたのか？

ここでは，$I(p)$ を定めるために log という関数を用いたことが恣意的なものではなく，自然なものであることを示そう．そのために，$I(p)$ の定義を決めるときに用いられた（はずの）原則について再考してみよう．

1. 「文」の情報量は「文」の価値を表すものと考える．「文」の価値が高いほど，その「文」の情報量も大きいとする．
2. 珍しい事象の生起を告げる「文」ほどその価値が高い．
3. 事象が珍しいほど，それを告げるためには長い「文」が必要になる．
4. 事象を伝えるために必要な「文」の長さの最小値をその「文」の価値とする．
5. 事象 A と事象 B が独立であるとき，事象 A と事象 B の両方が起きたことを告げる「文」は，概ね「A が起き，B も起きた」となるので，その長さの最小値は，事象 A が起きたことを告げる「文」（「A が起きた」）の長さの最小値と，事象 B が起きたことを告げる「文」（「B が起きた」）の長さの最小値の和である．
6. $I(p)$ の関数は p の微分可能関数にしておきたい．

第 5，6 項から $I(p)$ が $C\log_e p$ という関数であることが決まる．つまり，事象 A の生起確率を p，事象 B の生起確率を q とすると，事象 A と事象 B が独立であるとき，事象 A と事象 B の両方が起きる確率は $p\times q$ であるので，第 5 項は，

$$I(p\times q)=I(p)+I(q)$$

であることを要請している．このような条件を満足する微分可能な関数は，$I(p)=C\log_e p$ しかない（C は定数）．

【演習】上記を確かめてみよう．

補足

生起確率 p のメッセージのもつ情報量を単に $\dfrac{1}{p}$ にせず，それに log をつけて $\log_2\dfrac{1}{p}$ と定義する理由は，直観的には次のようにも考えられる．一般に，x が与えられたとき，$\log_2 x$ は x を 2 進数表示したときの桁数を表す．生起確率 p のメッセージの生起確率が 2^{-10} であったとき，そのメッセージを 2 進数を用いて表示するのであれば，どのような 2 進数表示を割り当てたらいいだろうか？

そのようなメッセージが他にもあり，それぞれに異なる 2 進数表示を割り当てなければ

ならないとすると，生起確率が2^{-10}のメッセージには10桁の2進数表示を割り当てるのが素直であるように思える．というのは，他のメッセージもすべて同じく生起確率が2^{-10}とすれば，そのようなメッセージは全部で2^{10}個あることになり，それらを区別するためには，10桁の2進数表示が必要であるからだ．他のメッセージのなかに生起確率が2^{-3}のメッセージがあるとすれば，それを表示するために，そのために10桁の2進数表示2^7個を確保して，その2^7個の10桁の2進数表示をひとまとめにして，$\log_2 \frac{2^{10}}{2^7} = 10 - 7 = 3$桁の2進数表示にしてしまう．このように，生起確率$p$のメッセージに長さ$\log_2 \frac{1}{p}$の2進数表示を割り当てるようにすると，すべてのメッセージに対する2進数表示の平均の桁数（＝長さ）は，そうでない割り当てをした時よりも必ず小さくなることが示される．つまり，各メッセージに対して異なる2進数表示を割り当てるという条件を課す限り，それよりも平均桁数を減らすことはできないという意味で，下限を与える．シャノンは確率モデルの下で発生するメッセージに対して有限個の記号を用いて異なる表示を割り当てたとき，その表示の平均長の下限を与える量として，情報量として規定したのである．このことはクラフトの不等式（第2回講義）を経て，情報源符号化定理（第4回講義）の中で示される．