E.T.ジェインズの「確率理論：科学の論理」の紹介（6）

原　宣一

ジェインズの「確率理論：科学の論理」の第1章と第2章までで、不確かな状況下で合理的で、常識と定性的に一致して、首尾一貫した、方法で推論すると必然的にラプラスの定義による確率になることを説明した。真実はベルヌーイが定義したものであるが、ラプラスの確率は無差別原理による確率と呼ばれるものであった。
一方、シャノンは不確かさのある状況がデータを得るとどのようにその不確かさが変わるかという論理的考察を進め、情報エントロピーの概念に到達した。そして確率は情報エントロピーが最大になるように決めなければならないことが導かれた。シャノンが考察の前提においた四つの条件はジェインズのデシダレータと本質的に同じであることが判る。確率の決め方は情報が全く無いときは無差別の原理によれば良いが、少しでも関連情報がある場合はエントロピー最大原理によることが一般的な原則になる。
ジェインズの「確率理論：科学の論理」第 11 章から、エントロピー最大原理が導かれる道筋を紹介する。エントロピーが出現する前に、やはり関数方程式が導かれる。

エントロピー：シャノンの定理
もし、確率分布によって表された「不確かさの量」の首尾一貫した測度があるならば、それが満たさなければならないある条件がある。

(1) ある数値的な測度$H_n(p_1,p_2,\cdots,p_n)$が存在すると仮定する。即ち、「不確かさの量」と実数の間にある種の関係を作ることが可能である。
(2) 連続性特性を仮定する。H_nはp_iの連続関数である。そうでないと確率分布の任意の小さな変化が不確かさの量において同じ大きな変化に導くからである。
(3) この測度が多くの可能性がある時に、少ないときよりも、より不確かであるという、常識に定性的に対応すべきであるということを要求する。この条件は p_i が全て等しい場合に次の形式を取る。その量、

$$h(n) = H_n\left(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n}\right) \qquad (11.6)$$

はnの単調増加関数である。これは「方向感覚」を確立する。
(4) 測度H_nが以前と同じ感覚で首尾一貫したものであるということを要求する。即ち、もし、その値を取り扱う一つ以上の方法があるならばあらゆる可能な方法に対して同じ答えを与えなければならない。

ロボットは二つの代替物を感知すると想定しよう。それにロボットは確率 p_1 と $q \equiv 1-p_1$ を割り当てる。そのときこの分布によって表された「不確かさの量」は $H_2(p_1,q)$ である。しかし、いまやロボットは第２の代替物が本当に二つの可能性から成るということを学び、それらに確率 p_2 と p_3 を $p_2+p_3=q$ が成立するように割り当てる。今や三つの不確かさ全部に関して彼の全体の不確かさ $H_3(p_1、p_2、p_3)$は何だろうか。さて、三つの内一つを選ぶ過程は二つのステップに分解され得る。最初に、最初の可能性が真かどうかを決めよう。この決定によって取り除かれる不確かさは元の $H_2(p_1,q)$である。それから確率 q で彼は事象２，３に関しての追加的な不確かさに遭遇し、次式を導く。

$$H_3(p_1,p_2,p_3)=H_2(p_1,q)+qH_2\left(\frac{p_2}{q},\frac{p_3}{q}\right) \qquad (11.7)$$

計算のどちらの方法によっても同じ正味の不確かさが得られる条件としてのものである。一般に、関数 H_n はこのような多くの方程式によってそれに関係する低次の関数に多くの異なった方法で分解できる。

(11.7)式は H_n が首尾一貫しているということを言っているだけでなく、それらが加法的であるということも言っている。

とにかく、次のステップは完全に率直な数学である。さて H_n に関する不必要な添字を落としてシャノンの定理全体を見てみよう。

最初に、考えるべき n 個の相互背反な命題（A_1、…、A_n）がある場合に対して、我々は構成則(11.7)の最も一般的な形式を見つける。それらに我々はそれぞれ確率(p_1、…、p_n）を割り当てる。（A_1、…、A_n）の確率を直接与える代わりに、我々はブール代数における（$A_1+A_2+\cdots+A_k$）によって表された命題として最初の k 個を第１番目のグループにし、(2-85)によって $w_1=(p_1+\cdots+p_k)$に等しい確率を与える。それから、次の m 個の命題は（$A_{k+1}+\cdots+A_{k+m}$）に結合され、それに対し確率 $w_2=(p_{k+1}+\cdots+p_{k+m})$を割り当てる。後、同様に。この多さが明示されたとき、複合命題に関しての不確かさの量は $H(w_1、\cdots、w_n)$である。

次に我々は複合命題（$A_1+\cdots+A_k$）が真であるということが与えられて、命題（A_1、…、A_k）の条件付き確率（p_1/w_1、…、p_k/w_1）を与える。確率 w_1 で遭遇した、追加的な不確かさはそのとき $H(p_1/w_1、\cdots、p_k/w_1)$である。複合命題（$A_{k+1}+\cdots+A_{k+m}$）、その他に対してこれを実行し、我々はあたかもその（p_1、…、p_n）が直接に与えられたかのような、究極的に同じ知識の状態に到達する。つまり、そのような首尾一貫性が、如何に選択がこの方法で破られていても、これらの計算で同じ究極の不確かさを生む

ということを要求する。こうして、我々は次式を持つ。

$$H(p_1 \cdots p_n) = H(w_1 \cdots w_r) + w_1 H(p_1 / w_1, \cdots, p_k / w_1) + w_2 H(p_{k+1} / w_2, \cdots, p_{k+m} / w_2) + \cdots \tag{11.8}$$

これは関数方程式(11.7)の一般形式である。例えば、

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right) \tag{11.9}$$

$H(p_1、\cdots、p_n)$は連続であるべきであるから、それは全ての合理的な次式の値に対してn_i個の整数でそれを決めるに十分だろう。

$$p_i = \frac{n_i}{\sum n_i} \tag{11.10}$$

しかし、そのとき(11-8)は関数Hを既にn個の等しく起こりそうな代替物の場合に対する不確かさの量を測るところのその量$h(n) \equiv H(1/n, \cdots, 1/n)$の項で決める。というのは、我々は次式の一つの選択において最初のステップとして代替物（A_1、…、A_n）の一つの選択として見なすことが出来るからである。

$$\sum_{i=1}^{n} n_i \tag{11.11}$$

ちょうど記述した方法で等しく起こりそうな代替物からの選択として、n_i代替物の間での選択でもある第２のステップも同様。一例として、n=3 で我々は $n_1=3$、$n_2=4$、$n_3=2$ を選んだかもしれない。この場合、構成則(11-8)は次式になる。

$$h(9) = H\left(\frac{3}{9}, \frac{4}{9}, \frac{2}{9}\right) + \frac{3}{9} h(3) + \frac{4}{9} h(4) + \frac{2}{9} h(2) \tag{11.12}$$

n_iの一般選択に対して、(11-8)は次式のように簡単になる。

$$h\left(\sum n_i\right) = H(p_i, \cdots, p_n) + \sum_i p_i h(n_i) \tag{11.13}$$

さて、我々はすべての $n_i = m$を選ぶことが出来る。つまり、ここで(11.13)式は次式のようになる。

$$h(mn) = h(m) + h(n) \tag{11.14}$$

明らかに、これは次式のように置くことによって解ける。

$$h(n) = K \log n \tag{11.15}$$

ここでKは定数である。しかし、この解はユニークなものだろうか。もし、m、nが連続変数ならば、これは答えるのが簡単である。つまり、mに関して微分して、m＝１とおき、そして結果としての微分方程式を(11.14)から明らかな初期条件 h(1)=0 で積分し、あなたは(11.15)がユニークな解であることを証明した。しかし、我々の場合で、(11.14)はただm、nの整数値に対してのみ成立することが必要である。

最初に、(11.15)はもはやユニークでないことに注目しよう。事実、(11.14)は整数mに対して無限の数の解を持つ。というのは、おのおの整数Nがユニークな構成を素数の要素で持つからである。従って、(11.14)の繰り返し適用によって、q_i が素数で mi が非負整数として、h(N)を $\sum_i m_i h(q_i)$ の形式に表現することが出来る。こうして $h(q_i)$ を素数 q_i に対して任意に明示することが出来る。ここで(11.14)が全ての正整数に対して h(n)を決めるに十分である。

h(n)に対してユニークな解を得るために、我々は h(n)が n において単調増加であるという我々の定性的要求を加えなければならない。これを示すために最初に(11.14)が帰納によって拡張されるかもしれないということに注目しよう。

$$h(nmr\cdots)=h(n)+h(m)+h(r)+\cdots \qquad (11.16)$$

そして、k 次の拡張において要素は等しいと置いて、次式が得られる。

$$h(n^k)=kh(n) \qquad (11.17)$$

さて、t 、s を2より小さくない何かの整数であるとする。そのとき、任意に大きなnに対して、次のような整数m を見つけるだろう。

$$\frac{m}{n}\le\frac{\log t}{\log s}<\frac{m+1}{n}、又は\quad s^m\le t^m<s^{m+1} \qquad (11.18)$$

h は単調増加なので、$h(s^m)\le h(t^n)\le h(s^{m+1})$ 。または(11.17)から、

$$mh(s)\le nh(t)\le(m+1)h(s) \qquad (11.19)$$

これは次式のように書ける。

$$\frac{m}{n}\le\frac{h(t)}{h(s)}\le\frac{m+1}{n} \qquad (11.20)$$

(11.18)、(11.20)を比較して、次のようであることが判る。

$$\left|\frac{h(t)}{h(s)}-\frac{\log t}{\log s}\right|\le\frac{1}{n}、又は\quad \left|\frac{h(t)}{\log t}-\frac{h(s)}{\log s}\right|\le\varepsilon \qquad (11.21)$$

ここで、

$$\varepsilon\equiv\frac{h(s)}{n\log t} \qquad (11.22)$$

は任意に小さい値である。このように、h(t)/log t は定数でなければならない。そして(11.15)のユニーク性は証明される。

さて(11.15)における異なったKの選択は対数を異なった基底に取ることと同じことになる。だからもし我々がその瞬間に任意の基底を残すならば我々はちょうど h(n)=log n と同じように書ける。これを(11.13)に置き換えて、我々はシャノンの定理を持つ。「不確かさの量」の合理的な測度に課した条件を満足するただ一つの関数 $H(p_1$、

…、p_n）は、

$$H(p_1,\cdots,p_n)=-\sum_{i=1}^{n}p_i\log p_i \tag{11.23}$$

この解釈を受け入れると、利用可能な情報によって課せられた制約に従う(11.23)を最大にする分布($p_1\cdots p_n$)はロボットが命題（A_1、…、A_n）について知っているところのものの「最も正直な」記述を表すであろうということになる。ただ一つの任意性は我々が望む対数の底を取ることが出来るということで、これはHにおける乗数に対応している。これは、もちろん、Hを最大にする(p_1、…、p_n)の値に影響を持たない。

我々は何を持ち何が証明されていないかの論理を注目する。測度(11.23)の使用が首尾一貫性のための必要条件である。しかしゲーデルの定理に従って、我々の証明に用いたものを越えた未知の領域に出てしまわない限り、人はそれが実際に首尾一貫しているか証明できない。元々はジェインズ（1957a）によって与えられ、シャノンに重く傾いている上述の論証から、我々は「情報測度」の他の選択はもし十分遠くに運ばれたなら不首尾一貫性に導くだろう。つまりこれの直接的証明は、後になって我々から全く独立な論証を用いてショアとジョンソン(1980)によって見つけられた。最大エントロピー原理の多年の使用は如何なる不首尾一貫性をも現わしていない。様々な筆者によって様々に省略され、PME、MEM、MENT、MAXENT などと表されている。そしてもちろん我々は別のものが見つけられることはないと信じている。

関数Hはエントロピーと呼ばれる。あるいは分布｛p_i｝の情報エントロピーと呼ぶ方が良いかもしれない。これはもう正せないように見える不幸な用語である。何らかの確率分布の特性である情報エントロピーと、定義されたように代わりに熱力学的状態の特性である熱力学の実験的エントロピー、例えばある物理系の圧力、体積、温度、磁化のような観察された量、の間を区別することの完璧な失敗である。これを始めに警告しなければならない。それらは決して同じ名前で呼ばれるべきでなかったのである。実験的エントロピーはいかなる確率分布に対しての言及もないし、情報エントロピーは何も熱力学の引用がない。しかし、問題がたまたま熱力学のものである場合には、それらの間に関係がある。多くの教科書や研究書は著者がこれらは全く異なったものであることを区別し損ねて、そして結果として、無意味な定理を証明していることによって致命的な欠陥を持っている。

幾つか前の章で、一般的に多項分布に関連して偶然に現れた数学的表現$\sum p\log p$を見ている。つまり今、それは一様確率分布が如何なるものかの基本的な測度としての新しい意味を獲得した。

上記の実証は数学的には満足出来るように現れているが、それはまだ概念的に完全に満足できる形式ではない。関数方程式(11.7)は我々の以前のものがそうであったようには直感的に優るものであるとは見えない。この場合、その困難は多分完全に確信させる方法で(11.7)を導く口頭による論証を我々はまだ学んでいないためだろう。多分これはあなたが自分の手で我々が (11.7)を書くことのちょうど以前に用いた冗長の多い言い回しを改善することを鼓舞するだろう。そのとき、同じ結論(11.23)をユニークに導く前に述べたショアとジョンソンのもののように５，６の他の可能な論証があるということを知ることは元気づける。

（６完）