少数の属性試験結果から得られる確信の度合い

原　宣一

成功か失敗かの二者択一の結果が得られる試験を属性試験という。付帯情報もなく、少ない標本数の属性試験結果のみ得られた品目の成功に対する確信の度合いを求める。確率の定義に確信の度合いとするものを採用し、情報が全くない事前の確信密度をラプラスの定義により与える。試験結果を見た後の事後確信密度をベイズの定理により求める。確信密度の期待値が求める確信の度合いに他ならず、これは標本数と成功数から簡単な式で求まる。

1．まえがき

観察できる実験回数の多寡に関わらず、いつの場合も常に着目する事象の生起について、その確からしさが知りたいことである。元来、確率の意味するものは確からしさという主観的な観念を数値化したものであり、心の問題であった［1］。

多くの試験結果を見れば確からしさを深めるものであるし、観察数が少なければ少ないなりに確かさの程度がある。このことを何らかの形で表現出来る筈である。人間の心の問題を客観的に扱うことが可能かについては異論もあろう。そこで、この議論を避けるためアンドロイド（人造人間）を想定し、彼に確信の度合いを問うことにする。このアンドロイドが普通の人間と違うところは、合理的に決められた規則に従って、主観的な観念を数値化して、即ち、元来の意味の確率で答えることにある。アンドロイドに組み込むべき確信の度合いを決める方法が必要である。

2．属性試験

ある種の試験は結果が合格か不合格かのどちらかしか判らない。このような試験は属性試験と呼ばれる。多くの火工品は１回の作動により消耗してしまうので作動確認試験は破壊試験であり、かつ属性試験である。

いま、製造元の信用とか過去の例というような関連情報が全くない火工品の成功確率を知りたいとする。この属性試験に供しうる個数が極めて少ない場合のこの品目に対する確信の度合いとしての確率を求める。

アンドロイドに与える情報は、試験個数nとその内の成功数rだけである。彼が合理的に算出する確信の度合いとしての確率はどのような式になるかを求める。

3．確信の度合いとしての確率

工学分野で通常採用されている確率の定義はフォン・ミーゼスによる頻度極限の概念である。この定義は客観的であるが、あくまでも母集団全体の特性値である。試験個数が少ないときに、この定義による確率は求めにくい。従って、確信の度合いと定義する確率を採用する。確信の度合い（Degree Of Belief）が主観確率として定義出来ることはサベイジ

宇宙開発事業団　安全・信頼性管理部
〒105-8060 東京都港区浜松町 2-4-1　　TEL 03-3438-6196　　FAX 03-5402-6515

により確立された［２］。確信の度合いとしての確率は心の状態を数値表現したものに他ならない。確率（ｐ:probability）とは「命題に対する確信の度合いである。」即ち、命題が真であることの確からしさについてその度合いを０から１までの数値で表現したものである。

このように定義した確率も、コルモゴロフの確率の公理を満たすように決めることが出来る。従って、すべての数学的確率論の定理が使える。ｐは０から１までの任意の値を与え得るが、次の３つの値は特別な場合である。

$p \mapsto 1$：命題が絶対的に真であるとの極めて強い確信を表す。（記号$p \mapsto 1$は数値１をｐに与える意味）

$p \mapsto 0$：命題が絶対的に偽であるとの極めて強い確信を表す。

$p \mapsto 0.5$：命題の真偽が全く不明の時の確信を表す。所謂、五分五分。

今、白石と黒石を入れた革袋から１個石を取り出す実験を考える。命題は「取り出した石は白である。」この命題が真であることに対して抱く確信の度合い、即ち確率（ｐ）はどのようなものかを問題とする。

革袋に白石だけが入れられた状況を見ていたら、$p \mapsto 1$である。黒石だけであったことを見ていたら$p \mapsto 0$である。黒30個と白70個を入れたことを知っていたら、$p \mapsto 0.7$である。白と黒を同数入れたことを知っていれば$p \mapsto 0.5$である。これらはラプラスの定義した確率で出された数値であり、また、この実験はｐの値についての強い確信がある場合である。

次に、白石と黒石のどちらをどれだけ入れたのか全く知らされていない場合を考える。このような状況の場合も、革袋から取り出した１個の石は白か黒かの二通りしか考えられない。本命題が真である確かさの度合い、即ち、確率（ｐ）は、$p \mapsto 0.5$となる。

白、黒同数の石を入れたのを見ていた場合も、同じ$p \mapsto 0.5$であった。両者の違いはｐに対する内容、つまりｐに対する確信密度が異なるのである。ｐに対する確信密度の分布$\pi(p)$の形状が異なるのである。

両者どちらの場合でも$p \mapsto 0.5$であるということをふまえ、「確率に対する確信密度の期待値はその確率に等しい。」とすることが適切である。確率は確信の度合いと定義することに伴う原理と捉えるべきものである。分布で表現した確率を代表値で表現するときの換算式として期待値を採用するということである。$p \mapsto E(p)$であるが、［$\mapsto$］を［＝］で書いてももはや混乱はない。

$$p = E(p) = \int_0^1 p\pi(p)dp \qquad \cdots(1)$$

前者は石を入れたのを見ていたという状況で、ｐの値について強い確信を持つことが出来る場合であり、その確信密度分布はディラックのデルタ関数で表すことが出来る（図１）。

白石と黒石を同数入れたことを見ている場合は、

$$\pi(p) = \delta(p-0.5) \qquad \cdots(2)$$

逆に、後者の場合は石をどのように入れたのか全く見ていないので、最も弱い確信の度合いの状態であり、その密度分布は０から１までの等分布で表すことができる（図２）。

$$\pi(p) = 1 \qquad \cdots\cdot(3)$$

後者の場合でも、石を１個、２個取り出して結果を見ることによりｐについての確信がだんだん強くなる。つまり、データを見ることによりｐの値に対する確信密度が変わるのである。

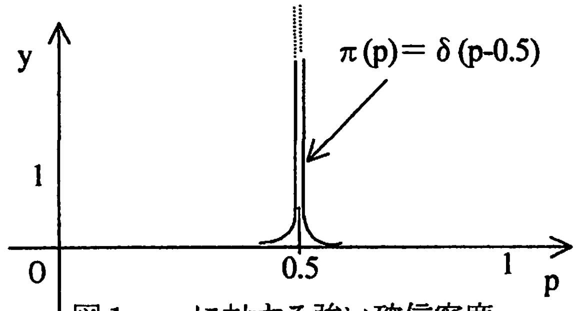

図１　ｐに対する強い確信密度

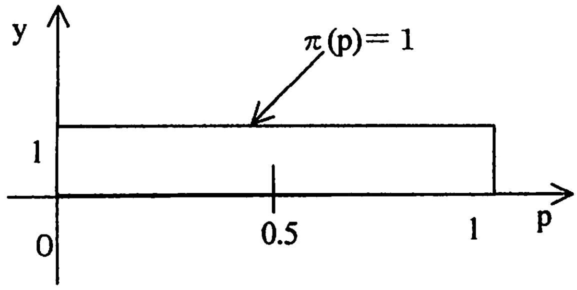

図２　ｐに対する弱い確信密度

確信密度π(p)は次のような性質を持っている。

$\pi(p)=0$、　$p<0$、$p>1$

$\pi(p)\geq 0$、　$0\leq p\leq 1$　・・・(4)

さらに、規格化の条件として次式を加えて良い。

$$\int_{-\infty}^{\infty}\pi(p)dp=\int_{0}^{1}\pi(p)dp=1 \quad \cdots(5)$$

図３は一般のｐに対する確信密度の例である。

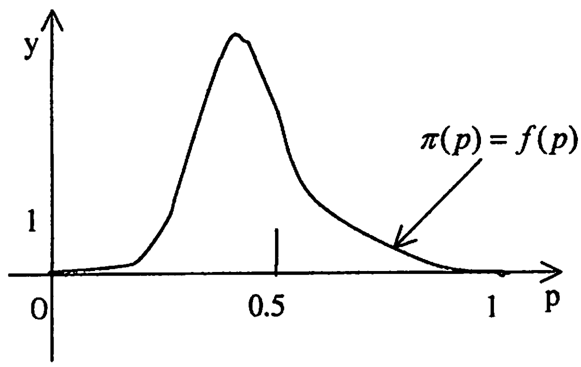

図３　一般のｐに対する確信密度

４．　ラプラスとダランベールの確率

先述の革袋から独立に２個の石を取り出す実験を考える。取り出した１個の石が白である確率をｐとすると、２個とも白である確率はp^2であるが、ｐについての確信密度がπ（ｐ）で表されるときは、（1）式が（6）式になる。

$$E(白、白)={}_2C_2\int_0^1 p^2\pi(p)dp \quad \cdots(6)$$

同様に、１個白で１個黒の場合、及び２個とも黒の場合は（7）、（8）式で表される。

$$E(白、黒)={}_2C_1\int_0^1 p(1-P)\pi(p)dp \quad \cdots(7)$$

$$E(黒、黒)={}_2C_0\int_0^1 (1-P)^2\pi(p)dp \quad \cdots(8)$$

（6）、（7）、（8）式のπ（ｐ）に（2）式を代入して計算すると、それぞれ 1/4、1/2、1/4 の値が得られる。これは２項分布Ｂ(2,0.5)と同じである。

一方、π（ｐ）に（3）式を代入して計算すると、それぞれ 1/3、1/3、1/3 の値が得られる。この結果は、取りうる状態が３種類あるのでそれぞれの状態に１／３の確率を割り当てた等確率分布になっている。

同時に投げた２枚のコインが２枚とも表、１枚表で１枚が裏、２枚とも裏である確率を求める問題で、ダランベールはそれぞれ 1/3、1/3、1/3 であるとした。ダランベールの弟子であったラプラスが、それぞれ 1/4、1/2、1/4 に正したと伝えられている。コインは１枚ごとに裏と表があるのでｐ＝0.5 の強い確信がある場合に相当するものである。実際に試行して統計を取ると相対比率はラプラスの与えた確率に近くなる事が認められている。

なお、上述のコインを独立にｎ個取り出す場合も同様な計算が出来て、２項分布になるか等確率分布になるかはπ（ｐ）に依存する。

５．　データを見た後の確信の度合い

ディラックのデルタ関数で表される程、事前の確信が強ければ、有限のデータを見ても確率は不変である。しかし、事前の確信が弱ければデータを見ることによって確信の度合い、即ち確率が変化する。この変化はベイズの定理により計算できる。

ベイズの定理はつぎのように主張している［2］。「データを見た後の事後密度はデータ

の尤度と事前密度の積に比例する。」

pに対する事後密度の期待値が求めるデータを見た後の確率pである。

属性試験は結果が成功か失敗かだけのデータを提供する試験であった。これは成功を白石とみることにより、革袋に入れる石を見ていない場合の先述の実験に相当する。皮袋の中は全部白石かも知れないが黒石がかなり混ざっているかもしれない。

まず、初めの状況は確信の度合いである確率pの事前密度は（3）式とすることが前述の考察により妥当である。

データ**X**は（x1、x2、・・・）は成功、不成功の列であるが、個々に独立としているので順序は関係せず、試験個数nと成功個数rだけが事後密度に影響する。

データ**X**を見た後の事後密度π(p｜**X**)は

$$\pi\left(p|X\right)\propto L(X|p)\pi(p)\quad\cdots(9)$$

ここでL(**X**｜p)はpに対するデータ**X**の尤度である。

試験個数nのうち成功個数rを得たというデータの尤度は$p^r(1-p)^{n-r}$であるから

$$\pi(p|X)\propto p^r(1-p)^{n-r}\times 1\quad\cdots(10)$$

（5）式の条件からベータ積分の公式を利用して定数を決めると、

$$\pi(p|X)=\frac{\Gamma(n+2)}{\Gamma(r+1)\Gamma(n-r+1)}p^r(1-p)^{n-r}\quad\cdots(11)$$

（11）式はβ-分布に他ならない。従って、データを見た後の確信の度合い、つまり確率pは（1）式により期待値を求めると、

$$p=E(\pi(p|X))=\frac{r+1}{n+2}\quad\cdots(12)$$

ここで、nは試験個数、rは成功数である。

（12）式が、データを見た後のアンドロイドが示す確信の度合いとしての確率値、又は信頼度である。

この式から、最初は信頼度0.5であったものが、失敗なしに8個連続の成功を見たあとで0.9まであがり、さらに続けて、失敗無しに18個の成功を見てやっと0．95まであがる。途中で1個でも失敗するとそのまま試験を続けたのではなかなか信頼度が回復しない事も判る。失敗原因の究明とその対策が必要なことを示すものである。

6．あとがき

本稿は昭和53年に宇宙開発事業団の社内研修会で「属性試験と信頼度」と題して発表したもの［3］を見直したものである。2，3の誤植を正し、説明を変えたところがあるが本筋は変わっていない。

属性試験は情報量が最小の試験である。試験データがパラメトリックに得られる場合は、目標値からの距離情報を上手く活かす方法が考えられるであろう。実験内容により多少の仮定が必要になる場合もあるが、標本数が少ない時は確信の度合いを確率と定義した議論の方が、頻度概念の確率を用いた議論よりは合理的であると考えられる。この一例を参考文献［4］に示す。

参考文献

［1］近藤次郎、応用確率論、日科技連、昭和45年

［2］D.V.リンドレー、確率統計入門1確率、竹内・新家共訳、培風館、昭和43年

［3］原、「属性試験と信頼度」、宇宙開発事業団社内発表前刷り集、昭和53年

［4］原、「疲労寿命とばらつき係数」、第14回構造強度に関する講演会前刷り集、昭和47年

（はら　のりかず）

表−1　試験個数と信頼度

n＼f	0	1	2	3	4	5	6	7
0	0.5							
1	0.667	0.333						
2	0.75	0.5	0.25					
3	0.8	0.6	0.4	0.2				
4	0.833	0.667	0.5	0.333	0.167			
5	0.857	0.714	0.571	0.429	0.286	0.143		
6	0.875	0.75	0.625	0.5	0.375	0.25	0.125	
7	0.889	0.778	0.667	0.556	0.444	0.333	0.222	0.111
8	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
9	0.909	0.818	0.727	0.636	0.545	0.455	0.364	0.273
10	0.917	0.833	0.75	0.667	0.583	0.5	0.417	0.333
11	0.923	0.846	0.769	0.692	0.615	0.538	0.462	0.385
12	0.929	0.857	0.786	0.714	0.643	0.571	0.5	0.429
13	0.933	0.867	0.8	0.733	0.667	0.6	0.533	0.467
14	0.938	0.875	0.813	0.75	0.688	0.625	0.563	0.5
15	0.941	0.882	0.824	0.765	0.706	0.647	0.588	0.529
16	0.944	0.889	0.833	0.778	0.722	0.667	0.611	0.556
17	0.947	0.895	0.842	0.789	0.737	0.684	0.632	0.579
18	0.95	0.9	0.85	0.8	0.75	0.7	0.65	0.6
19	0.952	0.905	0.857	0.81	0.762	0.714	0.667	0.619
20	0.955	0.909	0.864	0.818	0.773	0.727	0.682	0.636
	系列1	系列2	系列3	系列4	系列5	系列6	系列7	系列8

n：試験個数

f：失敗個数

図−4試験個数と信頼度