続　やさしい統計のお話 確率について)

（クォータリ-No.6 より）

前号で統計の基本的な考え方を述べました所，編集の方から引き続き投稿するように要請されましたので，今回は確率についてわかり易く書くことにします。

1.　確率とは

統計の基礎が「確率」にあることは既に皆さんお気づきでしょう。確率といいますと何となく面倒だな，と思う人も多いようですが，統計を勉強しようという以上はどうしても避けて通れないのです。確率とは，ある事象が起きるかどうかとか，存在するかどうかなどの「確からしさ」の度合いを，数字で示すものです。従って当たるも八卦，当たらぬも八卦のところがありますので，どうもスッキリしないなぁ，とよく指摘されます。例えば，将来のことで100%こうなる，などとはっきり予言できる筈はありませんから，当然なのですが，とかく人間は確定的なことを言って貰いたいという，欲張りな性格を持っていますから，何か釈然としない人もいるようです。

天気予報で「今日午前９時から午後３時までの，１mm以上の雨の降る確率は20%」などと放送します。今日はこの６時間の間に約1.2時間は雨が降り，約4.8時間は降らない，と解釈している人がいましたが，これは間違いです。「このような気象状況のもとでは，100回のうち20回の割合で１mm以上の雨が降る可能性があります。」という意味です。極端なことを言いますと，この６時間の間一滴の雨が降らなくても，あるいはずっとドシャブリでも，この予報は外れた，と非難されることはありません。

確率とは，そういうものであることを先ず認識しておいて下さい。それでも確率は極めて有用なのです。将来のこととか，サンプリングした僅かの試料から，ある母集団の姿を推定しようという場合にも，この確率の理論に頼らざるを得ないのです。

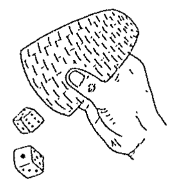
確率の話で，よく引き合いに出されるのはサイコロと宝くじです。６面体のサイコロをふって，例えば１の目が出る確率は１／６で，丁か半の出る確率は１／２であることはすぐわかります（ただしイカサマのサイコロではわかりません)。1,000万円を当てようと意気込んで宝くじを何回か買いましたが，全く期待外れでした。もし全部の宝くじを買い占めることが出来たら，必ず全ての賞金を一人占めすることが出来ます。ただし出資したお金の約半分は損をします。宝くじでは，賞金総額は売上額の約半分だそうです。

注　サイコロや宝くじの例は，確率が予めわかっています。これを先験的確率といいます。これに対し，われわれが実務で扱う母集団の推定などは，やって見ないとわかりません。これらを経験的確率といいます。

2.　サイコロをふる

サイコロの話に戻ります。１から６までのそれぞれの目が出る確率はみな同じで１／６ですから，全部の確率の合計は（１／６）×６で１となり，また丁半の何れかが出る確率は１／２ですから，全部の確率の合計は（１／２）×２＝１です。どんなに複雑な確率計算でもこの合計は必ず１となります。

注　１／６＝0.1666…，１／２＝0.5ですが，以下便宜上主として分数の形で扱うこととします。



さて今度は２つのサイコロを同時にふって，その２つの目の計を対象とすることとします。最小は１と１で計２，最大は６と６で 12 ですから，２から 12 までの数字が出ます。サイコロが１個のときは１から６まで全く同じ１／６ずつの確率でしたが，この場合は違います。計算して見ましょう。A，B２個のサイコロを使ったとして，例えば計が２になるのは（A＝１，B＝１）のときですから，A＝１の出る確率は１／６，B＝１も同じく１／６，これらが同時に起きる確率は，掛け算で（１／６）×（１／６）＝（１／36）となります。計が３になるのは（A＝１，B＝２）のときと（A＝２，B＝１）の２つの組合せがあり，１つの組合わせにつき，確率は１／36ですから確率の合計は２／36となります。このようにして，２つのサイコロの

表１　２個のサイコロの計の確率

（確率の合計　36／36＝１）

２個の計	計の出る組合せ (A, B)	組合せの数	起きる確率
2	(1,1)	1	1/36
3	(1,2) (2,1)	2	2/36
4	(1,3) (2,2) (3,1)	3	3/36
5	(1,4) (2,3) (3,2) (4,1)	4	4/36
6	(1,5) (2,4) (3,3) (4,2) (5,1)	5	5/36
7	(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)	6	6/36
8	(2,6) (3,5) (4,4) (5,3) (6,2)	5	5/36
9	(3,6) (4,5) (5,4) (6,3)	4	4/36
10	(4,6) (5,5) (6,4)	3	3/36
11	(5,6) (6,5)	2	2/36
12	(6,6)	1	1/36

表２　３個のサイコロの計の確率

（確率の合計　216／216＝１）

３個の計	組合せの数	起きる確率	３個の計	組合せの数	起きる確率
3	1	1/216	11	27	27/216
4	3	3/216	12	25	25/216
5	6	6/216	13	21	21/216
6	10	10/216	14	15	15/216
7	15	15/216	15	10	10/216
8	21	21/216	16	6	6/216
9	25	25/216	17	3	3/216
10	27	27/216	18	1	1/216

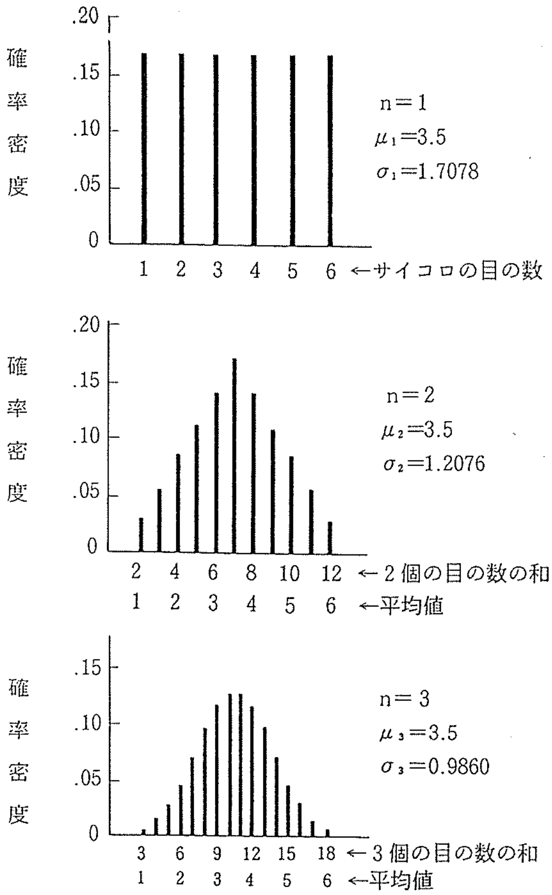


図１　確率分布のグラフ

目の計が，２から12まで出る確率を全部計算してまとめたのが，表１です。

今度は３個のサイコロをふることにします。３個の計は３から18までになります。前と同じように組合せの数と確率の計算をしてまとめたものが表２です。

さて，以上の試行は，サイコロが１個のとき，さらに２，３個のときは計をとってそれぞれの確率を計算しました。さてもう少し進展させて，ｎ＝１，２，３とし，計をｎで割った値（つまり平均値）の分布を考え，それぞれの分布の平均値と標準偏差を計算して見ましょう。

注　標準偏差については，他の参考書に譲ることとしまして，ここではただ「バラツキを示す尺度」とだけ説明しておきます。しかもこの場合は，私たちがよく経験している，試料からのデータについて平均値と標準偏差を計算するのとは，式が違っていますが本質は全く同じですから余り気にしないで，結果だけに注目して頂いて結構です。一応計算式を書いておきます。

母平均：$\mu=\Sigma x \cdot P(x)$

母分散：$V=\Sigma\{(x-\mu)^2 \cdot P(x)\}$

母標準偏差：$\sigma=\sqrt{V}$

$P(x)$：ｘがとる確率

図１はｎ＝１，２，３の場合の平均値（勿論ｎ＝１ではもとの数値）の確率の分布状態を示したものです。また，計算で得られた母平均と母標準偏差も付記しておきました。縦軸の確率密度とは，要するに確率の数字で，合計して１になるようになっています。

これだけの計算ですが，次のような極めて重要な理論がわかります。

① 総平均値は，ｎに関係無くみな同じである。（母平均）

② グラフからもわかるように，ｎ＝１のとき（元の分布）は全く矩形分布であったのに，平均値の分布をとるとｎ＝２，３と増えるに従って，総平均値のあたりが最高で，すそを引いている「つりがね」型に近づいて来る（正規分布に近づく）。

③ 標準偏差を比較して見ると，nが多いほど小さくなっている。即ちバラツキが減っている。しかも明瞭な数量的な関係がある。n個のときの平均値の標準偏差は，元の標準偏差 σ_1 の $1/\sqrt{n}$ である。n＝2，3のときの標準偏差をそれぞれ σ_2 および σ_3 とすれば

$$\sigma_1/\sqrt{2}=1.7078/1.4142=1.2076\ (=\sigma_2)$$
$$\sigma_1/\sqrt{3}=1.7078/1.7321=0.9860\ (=\sigma_3)$$

以上の性質は、元の分布がどんな形をしていても成立します。この場合は元の分布は矩形です。一般にnを4か5にして平均値をとり，その平均値の分布を考えると，元の分布がどんなに歪んでいても，あるいは元の分布の形がわからなくても，正規分布の理論を適用出来ることになります。統計を応用する上での基本の一つです。

3. ランダムということについて

確率の理論，即ち統計の理論が成立するための，大前提条件があります。サイコロがイカサマであったり，インチキなふり方ではこれまでの理論は全く崩れてしまいます。いわゆるランダム（無作為）であることが必要です。ある辞書によると，ランダムとは「手あたり次第に」とありましたが，少しニュアンスが違うようです。皆さんの仕事でのサンプリングで，良いものをとか，特に悪そうなところから，などの意識を働かせてはならないのは，データ処理の上で統計理論が全く適用出来なくなり，母集団の姿を間違って判断する恐れがあるからです。よく過去のデータを大量に持ち込んで，統計解析を，と依頼されることがありますが，第一にそれらのデータがランダムにとられたものかどうかを，十分に確認する必要があります。ランダム（無作為）に，とは「母集団を構成する単位（物，量，事象など）が，何れも同じ確率でとられるように」という意味です。場合によっては乱数表を使うなど，積極的な無作為化を計ることも必要となります。

4. むすび

本号の内容は少し理屈が多くて，皆さんにとって馴染みにくかったかもしれません。しかしそう難しいことではありませんから，是非理解して頂きたいと思います。出来れば，子供さんを相手にして，サイコロをふってデータをとり計算して，ここに書いた理論をトレースして見ることをお勧めします。試みの数は，多ければ多いほど，理論値に近づきます。