

研究員 の眼

世の中の数字の現われ方は一律ではない って知っていましたか —ベンフォードの法則について—

取締役 保険研究部 研究理事

年金総合リサーチセンター長

TEL: (03)3512-1777

中村 亮一

E-mail: nryoichi@nli-research.co.jp

はじめに

「対数 (logarithm)」という言葉を知ると、何となく身構えてしまう人が多いのではないかと。学生時代に、数学で学んで、試験問題でも苦労した人もいるかもしれない。そもそも、日常生活においては、通常、一般の人が「対数」に出会うことはないと思われる。私も大学卒業後、確率や統計等を扱う仕事に従事してきているが、入社後、専門的な資格試験をパスするために、多くの試験問題を解くのに苦労して以来、日常業務で「対数」を使うことは殆どなかった、と思われる。

ただ、実は、自然界には、この「対数」に基づいている測定結果が幅広く存在している。「ベンフォードの法則 (Benford's Law)¹」は、こうした自然界における測定結果の最初の桁の数値の分布が、一様ではなく、対数を用いて表される特定の分布に従っているというものである。

数字の出現確率

我々が、自然界での各種事象に現われてくる数字を観察してみた場合、1から9までの数字が一様に出現してくるわけではない。直感的にも一定程度納得できると思うが、実は1とか2とか若い数字ほど出現確率が高くなっていることが多い。

即ち、一定の事象については、ベンフォードの法則と呼ばれるものに従って、最初の桁の数値の分布は、以下の通りとなることが知られている。この出現確率は、対数を用いて、 $Y = \log_{10}(1 + 1/X)$ で表される。

最初の桁の数字(X)	1	2	3	4	5	6	7	8	9
出現確率(Y)	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%

これによれば、最初の数字の約半分が1か2ということになる。

¹ 1938年にこの法則を提唱した物理学者フランク・ベンフォードに因んで名付けられている。

ベンフォードの法則の適用例

ベンフォードの法則は、電気・ガス・水道等の公共料金の請求書、株価、物理・数学定数、川の面積、スポーツの成績、さらには人口や新聞記事等にも適用できるとされている。

ただし、全てのケースに適用できるわけではない。ベンフォードの法則が成り立つためには、「尺度不変性」（0以外の数値をかけても分布が変わらない）ことが必要とされている。従って、例えば、数値の範囲が制限されていたり、別のルールに基づいて定められている数値（電話番号、自動車のナンバープレート、各種の証明書番号等）には適用されない。

株価の出現確率

何故このようなことになるのかを、株価を例にとって考えてみる。

1つの想定として、「株価が2倍になっていく時間が一定」ということを考える。即ち、1年間で株価が2倍になるのであれば、その後の1年間でさらに2倍になる、こととする。日本における現在の市場環境を踏まえると、あまり現実味は無いが、長期的に持続的な成長が期待できる環境や市場で、平均的な株価の成長を考えれば、一定程度は納得できるのではないかと考えられる。

さて、この場合のスタート時点で1の株価がどのように推移していくのか及びその結果として株価の最初の数値がどのような確率で現われるのかをみてみると、以下の通りになる。

株価が10に到達した後は、1から10への成長と同じことが繰り返されることになる。

株価の推移と数値の現われ方

① 経過(年)	0	1	1.585	2	2.322	2.585	2.807	3	3.17	3.322
② 株価	1	2	3	4	5	6	7	8	9	10
③ 1桁目がそれぞれの数値である期間	1	0.585	0.415	0.322	0.263	0.222	0.193	0.17	0.152	-
④ 1桁目がそれぞれの数値である確率 (③÷3.322)	30.1%	17.6%	12.5%	9.7%	7.9%	6.7%	5.8%	5.1%	4.6%	-

この表の④がまさに1ページ目の出現確率に一致することになる。

このことは、数学的にも簡単に説明できる。

株価Xに到達する経過をa、株価X+1に到達する経過をbとすれば、

$$2^a = X \quad 2^b = X+1$$

$$a = \log_2 X \quad b = \log_2 (X+1)$$

$$b - a = \log_2 (X+1) / X$$

株価が丁度10になる時間が $\log_2 10 (=3.322)$ となるので、株価の最初の数値がaである確率は

$$(b - a) / \log_2 10 = (\log_2 (X+1) / X) / \log_2 10 = \log_{10} (1 + 1/X)$$

ということになる。

自然界での事象

自然界では、ある事象が発展していくとき、一定の時間間隔で等比（同じ比率）的に成長していくことが考えられる。具体的には、例えば、細菌の増殖等のように、一定の時間間隔で2倍に増えていく（倍々になっていく）ことが往々にして見られると考えられる。

こうした事象については、ベンフォードの法則があてはまることになる。こうして考えると、ベンフォードの法則も自然なものだと思えてくるのではないか。

ベンフォードの法則の応用

ベンフォードの法則は、会計等における不正のチェックの検証に使用することができる。

一般的に不正を行うために、データを作為的に作成する人は、できる限り普遍的に数値を分布させようとするが、これは理論的には却って不自然な結果をもたらすことになるからである。

例えば、会計監査における不正経理の検出や各種の科学的・統計的なデータの検証等に応用できることになる。

最初の桁以降への拡張

ベンフォードの法則は、最初の桁以降にも拡張できる。

ある複数桁を有する特定の数値 P に遭遇する確率は、上で説明したのと同じ考え方から $\log_{10}(1 + 1/P)$ となる。具体的に、例えば、数字の最初の 2 桁が 77 で始まる確率は、 $\log_{10}(1 + 1/77)$ (≈ 0.0056) となる。

同様に、数字の特定の桁に、ある数字が現われる確率も求めることができる。

例えば、2 桁目に 7 が現われる確率は

$$\log_{10}(1 + 1/17) + \log_{10}(1 + 1/27) + \dots + \log_{10}(1 + 1/97) \approx 0.09035$$

となる。一方で、2 桁目に 1 が現われる確率は

$$\log_{10}(1 + 1/11) + \log_{10}(1 + 1/21) + \dots + \log_{10}(1 + 1/91) \approx 0.11389$$

となる。

これからわかるように、 n 桁目の数字の分布は、 n が増加するにつれて、一様分布に基づく出現確率である 10% に近づいていくことになる。これは、一般的な感覚とも一致していると思われる。

なお、実際の不正目的の発見においては、こうした 2 桁目以降の確率も用いられている。

まとめ

世の中には、思わぬところに思わぬ法則が隠れていて、社会にも役立っている。自然界に現われている事象は、多様性があり、結構ランダムに行われているようで、実は何らかのルールに基づいているケースも多いということであり、大変興味深いことだと思われる。

併せて、これを機会に、馴染みが薄い「対数」にも少しは関心を持ってもらえればと感じた次第である。