研究員の眼

ランダムかどうかの判断

分析に用いている乱数は、本当にランダムといえるか?

保険研究部 主任研究員　篠原　拓也
(03)3512-1823　tshino@nli-research.co.jp

確率論や数理統計に基づいて、統計分析をしたり、数理モデルを構築したりする際、しばしば「ランダム（無作為）」かどうか、が問題となる。標本調査では、標本をランダムに抽出する必要がある。為替や株価の時系列分析をするときには、観測データにランダムな変動が含まれることを想定している。

例えば、経済予測をする際に、経済モデルを設定して、シミュレーションを行うことがある。そこで用いる金利や為替等の前提条件に、ランダムな変動を織り込むことで、シミュレーション結果の現実味を高めることができる。このため、昔から、計算によってランダムな数（乱数）を発生させる手法が考えられてきた。

しかし、計算式で乱数を発生させようとしても、最初はそれらしいものが得られるが、長い数列を見ていくと、やがて同じ数が繰り返し現れてしまう。そこで、数列の長さをものすごく長くして、簡単には繰り返しが起こらないようにするといった工夫が考えられている。しかし、それでも、何億個、何兆個もの乱数列を計算式で発生させようとすると、繰り返しの問題が生じてしまう。このように、ある計算式によって得られる数列は、本当の乱数ではないので、擬似乱数と呼ばれている。そもそもランダムである乱数を、何かの計算式を使って発生させるというのは、自己矛盾していると言えよう。

そこで、計算式で発生させる代わりに、既にある数を利用してはどうか、という発想に行き着く。その代表的なものとして、円周率πが挙げられる。
π＝3. 141592653589793…　と続く円周率は、小数点以下が無限に続くことが知られている。それだけではなく、いまのところ、小数点以下の数値の並びに規則性が見つかっておらず、その一部分を切り出して乱数として使うことができる。ただし、何か隠れた数学的な規則性があるに違いない、と信じている数学者や統計学者も多く、その研究も進められている。円周率の数の並びは、現在知られている中で、かなりランダムに近いと言えよう。

次に、与えられたものが、ランダムかどうかを判断することを考えてみよう。例えば、a と b を 20 個並べた文字列である、列Ⅰと、列Ⅱが、与えられたとしよう。

順番	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
列Ⅰ	a,	a,	a,	a,	a,	b,	b,	a,	b,	b,	a,	a,	b,	a,	b,	b,	a,	b,	b,	b
列Ⅱ	a,	b,	b,	a,	b,	a,	b,	a,	a,	b,	a,	a,	a,	b,	a,	b,	b,	a,	b,	b

列Ⅰでは、早い順番の方に a が多い。一方、列Ⅱには、そのような傾向は見られない。感覚的に、列Ⅰと、列Ⅱでは、ランダムかどうかに、違いがあるように見える。

このように 2 種類の文字が並んだ列のランダム性を、統計的に判定する手段として、「ウィルコクソンの順位和検定」という方法がある。この方法は、文字の列がランダムであるかどうかの判定ではなく、ランダムではないかどうかの判定を行う。以下に示すとおり、具体的な内容は、かなり技術的だが、感覚ではなく、数量的に、ランダム性についての判定を行うことができる。

列Ⅰを、例にとって、判定を行ってみる。

【ステップ 1】a と b のうち、個数が少ない方について、順番の数を足す。個数が同じ場合は、どちらについて足してもよい。

例では、a も b も 10 個ずつで同数のため、a について足すことにする。

1 ＋ 2 ＋ 3 ＋ 4 ＋ 5 ＋ 8 ＋ 11 ＋ 12 ＋ 14 ＋ 17 ＝ 77　となり、順番の数の和は、77 となる。

【ステップ 2】個数が少ない方の文字の個数と、全体の個数に 1 を足した数とを乗じて、そこからステップ 1 の数を差し引く。

例では、a が 10 個、a と b の全体で 20 個あるので、10 ×（20 ＋ 1）－ 77 ＝ 133　となる。

【ステップ 3】ステップ 1 と、ステップ 2 で得られた数のうち、小さい方の数をとる。

例では、ステップ 1 の 77 と、ステップ 2 の 133 のうち、小さい方として、77 となる。

【ステップ 4】ステップ 3 の数が、統計学であらかじめ算定される、所定の数よりも小さい場合、「ランダムではない」と判定する。ステップ 3 の数が、所定の数以上の場合には、「ランダムでないとは言い切れない」こととなる。所定の数は、判定の信頼度により異なる。ここでは、信頼度を 95%、つまり 20 回に 1 回程度しか判定を誤らない水準と置く。a と b が 10 個ずつの場合、所定の数は 82 となる。

例では、ステップ 3 の数 77 が、所定の数 82 を、下回っている。

従って、列Ⅰは、ランダムではないと判定される。

なお、列Ⅱで同様の判定を行うと、ステップ 3 の数は 97 となり、所定の数 82 を上回る。

従って、列Ⅱは、ランダムでないとは言い切れない。

以上のように、乱数を作ることは難しい。また、乱数が、本当にランダムかどうかを判断するのにも、手間がかかる。しかし、シミュレーションでは、乱数を用いて調査や分析を行うことで、結果の現実味を高めることができる。シミュレーションの結果を見る際は、乱数の設定方法にも注意をしてみてはどうかと思われるが、いかがだろうか。