

確率と統計の違い

「確率とは何ですか」といったときによく返ってくる返事は「確からしさを数字にしたものです」というものです。例えば「サイコロを振った時に1～6の目が出る確率は等しい」とか「コインをたくさん投げると表と裏の出る確率は半々である」とか言われます。これは常識的には正しいと思われませんが、学問としての確率をいうときにはこの答は確率と統計を混同していて正確ではありません。たとえば1が多く出るサイコロもあるし、表のほうが多く出るコインがあってもいいのです。サイコロやコインが歪んでいればこのようなことは起こり得ますが、それとは関係なしに表が出る確率が1、裏が出る確率が0などという確率があってもいいのです。

ある大学の数学科の教授が学生に「確率とは何か」と質問したそうですが、そのときもサイコロやコインがどうのこうのという答えをした学生がいたそうです。一般人ならともかく、理系の学生がこの答を出したとしたら確率を全く理解していないということになり、教科書を最初から読み直せと言われても仕方ありません。

数学では確率論は測度論といわれるものの1つなのですが、これは理系の大学生でも単位を落とすことが多い専門的な科目なので触れないとして、数学的には確率は対象となる事柄を0から1までの間の数値に対応づけるという作業（演算）であり、その作業は以下の規則を満たすものとされます。

1. 起こり得るすべての事柄に対応する確率は1
2. 起こりえない事柄に対応する確率は0
3. 2つの事柄が同時に起こることがない場合、2つの事柄のどちらかが起こる確率はそれぞれが起こる確率を足したものとなる。

これを確率の公理といい、これを満たすような対応づけを確率といいます。

従って確率はきわめて形式的に作ることができ、コイン投げに関していえば表が出る確率が100%で裏が0%などという確率も決めることができます。これは常に表が出続けるコインということになり、こういうコインが存在するかどうかは統計の問題となるのです。統計というのは過去の実際の事柄に関する数値を集めたものであり、それを使って将来起こることを予想したりするのに使われます。また予想がどのくらいの精度で正しいかという計算もできます。そのためにその統計数値の平均をとったり、分散（ばらけ具合）を計算したり、2つのことがらの間の関係（相関）を調べたりといろいろな作業をするのですが、その作業で使われるのは確率論により作られたさまざまな定理や公式なのです。つまり確率は統計を扱うための数学の道具であり、それ自体は実際の出来事とは関係ないものなのです。