



プログラムで理解する視覚でわかる統計確率論 第3回

ベイズの定理で確率を考える

浪平 博人

あるデータが得られたときに、その結果を反映した後の事後確率を求めるのに使われている定理があります。この考え方は推論などで使われ、近年では社会科学や経済学、コンピュータの世界ではベイジング・フィルタとしてスパム・メール対策などにも広く利用されています。このベイズの定理について理解を深めます。(筆者)

ベイズの定理は、ある出来事を観察することで事象の起こりやすさが観察前より変わることを主張し、その変化の度合いを計算する公式を提供するものです。この方法は、スパム・メールを見つけるベイジアン・フィルタとしても利用されています。今回はこのベイズ統計について解説します。

条件付き確率

最初に、条件付き確率の話から始めましょう。

サイコロの目が出る確率はどれも同じで、確率1/6で起きます。たとえば2の目が出る確率は1/6です。しかし、目が偶数である(2, 4, 6のどれか)という条件をあらかじめ設定すれば、2の目が出る確率は1/3となります。このように、ある事象Xが成り立っているという条件のもとに事象Yが起きる確率を条件付き確率と呼び、 $P(Y|X)$ と記します(図1)。

$$P(Y|X) : \text{条件} X \text{のもとで} Y \text{が起きる確率} \\ (\text{条件付き確率}) \dots\dots\dots (1)$$

これは、次のようにも書けます。

$$P(Y|X) = P(X \text{ and } Y) / P(X) \dots\dots\dots (2)$$

● つぼからコインを取り出す例題

次に、条件付き確率の例題を考えてみましょう。

正しいコイン(以下、正貨)2枚の入ったつぼAと、正貨と偽のコイン1枚ずつ入ったつぼBがあるとします。つぼ

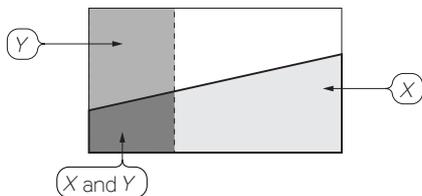


図1 事象Xと事象Yが起きる確率の模式図

A, Bをランダムに選び、選んだつぼからコインを無作為に取り出し、元に戻します(図2)。

取り出したコインが正貨であるとき、選んだつぼがAである条件付き確率は次のようになります。ここで、取り出したコインが正貨である事象を記号X, 取り出したコインが正貨である確率を $P(X)$ とします。

$$P(X) = A \text{ を選ぶ確率} \times A \text{ から正貨を取り出す確率} \\ + B \text{ を選ぶ確率} \times B \text{ から正貨を取り出す確率} \\ = 1/2 \times 1 + 1/2 \times 1/2 = 3/4 \dots\dots\dots (3)$$

$$\text{条件付き確率: 取り出したコインが正貨のとき、} \\ \text{選んだつぼがAである確率} \\ = P(A | X) = P(A \text{ and } X) / P(X) \\ = (1/2) / (3/4) = 2/3 \dots\dots\dots (4)$$

これにより、1回の取り出しで正貨が出たとしても、つぼが正貨のみが入ったAであるとは強く主張できません。Bである可能性が1/3はあるのですから。

しかし、選んだつぼからコインの取り出しを2度行い、両方とも正貨であったという条件のもとでは、そのつぼが正しいつぼAである確率は高くなりそうです。

2度とも正貨である事象をXXとしましょう。XXはつ

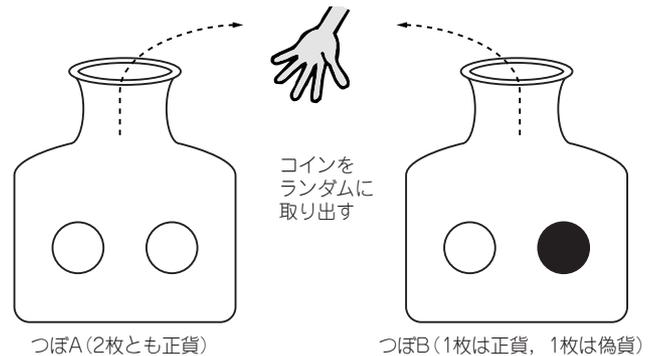


図2 正貨と偽貨が入ったつぼからランダムにコインを取り出す