確率論の基礎

結合確率と事後確率

用語の定義

試行(experiment) ¶

繰り返し行え、偶然に支配されるような実験や観測を試行という。

> 「今日の天気を観測する」

母集団(universe)

調査対象となる集合全体。

> 「過去100年、未来100年の 岡山の天気データ」

標本(sample)

母集団の部分集合。

> 「今年の秋の岡山の天気データ」

標本化(sampling)

標本を作成すること。

> 「今年の秋の岡山の天気データを収集する」

事象(event)

できごと。標本空間の部分集合のうち特別に選ばれたもの。正か誤、真か偽、あるいは複数の可能な状態のいずれかであることが明確に区別できる必要がある。確率で表わす対象を、確率事象、あるいは単に事象(event)と呼ぶ。

> 「今日、岡山に(1mm/h以上の)雨が降る」

余事象(complementary event)

標本空間の部分集合のうち、事象に含まれないものすべて。

- ○「今日、岡山に雨が降らない」これは余事象
- ×「今日、倉敷に雨が降る」これは余事象ではない。

事象xに対する確率を$P(x)$と表すことにする。事象xと事象yが同時に起こる確率を、$P(x, y)$と書き、xとyの結合確率あるいは同時確率(Joint probability)と呼ぶ。

事象は集合である。xを、「サイコロを一回ふって1が出る」事象とし、yを「サイコロを一回ふって2が出る」事象とすれば、$P(x) = P(y) = 1/6, P(x, y) = 0$となる。また、必ず1から6の目のどれかが出るので、確率の総和は1になる。

$$\sum_{i=1}^{6} P(i) = 1$$

サイコロの目のとりうる状態全体をAと表わす場合もある。

$$\sum_{A} P(A) = 1$$

と書いてある場合は、

$$\sum_{x \in A} P(x) = 1$$

と読みかえて欲しい 。

条件付き確率、あるいは事後確率(posterior probability)は$P(x|y)$と書く。事象yが起こったという条件のもとで、事象xが起こる確率のこと。縦棒は "when" あるいは「ただし」と読めばいい。

事後確率という言葉がわかりにくいかもしれない。$P(x|y)$は、yのあとにxが起こる確率ではない。yが起こったことを知った上で、xが起こる(起こった)確率のこと。事後というのは、観測者がその事実を知った順番を表していて、できごとx, yの時間的な順番を表しているわけではないことに十分注意してほしい。

xのとりうるすべての状態に関する、$P(x|y)$の和は1になる。

$$\sum_{x} P(x|y) = 1$$

事象y, zがともに起こったという条件のもとで、事象xが起こる確率は、$P(x|y, z)$と書く。これは$P(x|(y \cap z))$と同じ意味だが$P((x|y), z)$ではない。(後者は$P(x, z|y)$と書くべき)

事象xとyが、次の関係をみたす時、**独立事象**という。

$$P(x) = P(x|y)$$

確率論の基本定理

2つの基本定理を知っておけば十分。

加法定理

$$P(x) = \sum_y P(x, y)$$

乗法定理

$$P(x, y) = P(x|y)P(y)$$

---

xとyが独立事象なら、$P(x) = P(x|y)$なので、$P(x, y) = P(x)P(y)$である。

もっとパラメータが多い場合にも、上の定理が成りたつ。例えば、
$P(x, y, z) = P(x|y, z)P(y, z) = P(x|y, z)P(y|z)P(z)$ $P(x|y, z) \ = P(x, y, z)/P(y, z)$ $P(x, y|z) \ = P(x, y, z)/$

乗法定理によれば、$P(x, y)$は2通りに表すことができる。

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

これより、ただちに次のBayesの公式が導かれる。

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

Bayesの公式は次のように使える。たとえばyを「西の山に雲がかかる」事象とし、xを「翌日に雨が降る」事象とする。$P(x|y)$は西の山に雲がかかった(のを知った)場合に、翌日雨が降る確率。これに対し、$P(y|x)$は雨が降った前日に西の山に雲がかかっていた確率である。一年を通して、雨が降る確率$P(x)$、西の山に雲がかかる確率$P(y)$、そして雨が降った前日に西の山に雲がかかっていた確率$P(y|x)$を観測しておけば、$P(x|y)$がわかる、という仕掛けである。天気予報に使える情報源がほかにもあるとしよう。西の山の雲だけでなく、いろいろ関係ありそうな情報(カエルが鳴いた、月にカサがかかった、など)を事象a, b, c, dと表すと、

$$P(x|a, b, c, d)P(a, b, c, d) = P(a, b, c, d|x)P(x)$$

事象a, b, c, dが全部起こる確率$P(a, b, c, d)$と、$P(x)$と、雨が降った前日に事象a, b, c, dが起こっていた確率$P(a, b, c, d|x)$の統計をとっておけば、事象a, b, c, dが全部起こった時に、翌日雨が降る確率を求めることができる。統計的な裏付けをもって、未来を高い確率で予測できる。原因から結果が正確に予測できる。つまり確率的な因果関係(のようなもの)を知ることができる。

何が原因かはよくわからないが、核生成のように突然起こる事象があるとしよう。その事象が起こる確率は、その周辺の様々な量の統計と、事象が起こる直前での統計を見比べることで、計算できる。周辺の状況がどうであれば、どれだけ核生成確率が上がるかを数値で示せる。

```
In [ ]:
```