第１０４回数学教育実践研究会　レポート発表

ベイズ統計学を用いた数学Ａでの条件付き確率の導入の工夫

北海道札幌手稲高等学校　　西村　昂介

平成３０年１月２７日（土）　ニッセイ MK ビル

１．はじめに

現行の数学Ａの確率の単元に条件付き確率が新たに加わったが、どうも生徒もわかりにくさを感じているし、試験の中でも条件付き確率という言葉を見ると難しい問題ととらえてしまう生徒が多いように感じている。

日ごろから、この「条件付き確率」をよりわかりやすく、実生活との結びつけて教えたいと思っていた中で出合った本を今日は紹介したいと思います。

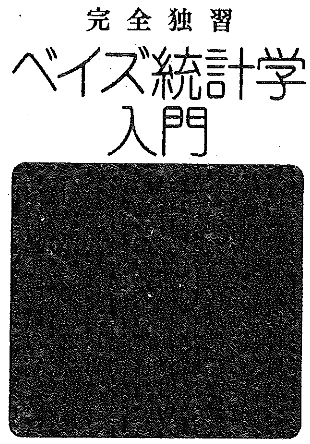
また、もうすぐ、バレンタインデーが近いので、男子としては、チョコをもらった場合にそれが、どれくらいの確率で本命チョコなのかを、この本の一部を紹介しながら考えていきたいと思います。

先生方の授業で話題としてもらえたらありがたいです。残念ながら、私は現在、１年生を教えていないため、話題にあげられず、悲しい思いをしています。

２．書籍の紹介と具体例

小島博之著

「完全独習　ベイズ統計学入門」です

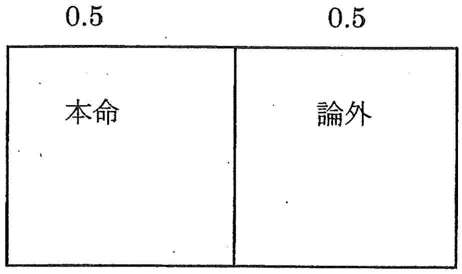


この中の第３講にある問題を紹介します。

問題

> あなたが男性であると仮定し、特定の同僚女性さんが自分に好意を持っているかどうか気になっているとする。そんな中、あなたはバレンタインデーに彼女からチョコレートをもらった。さて、あなたは、彼女が自分を本命と考えている確率をいくつと推定すべきか。

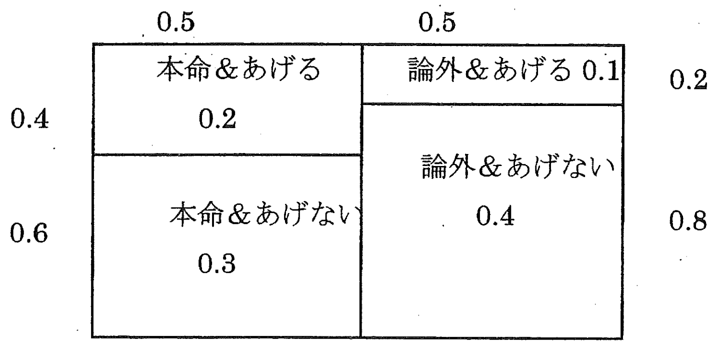
まず、事前確率を設定します。
相手が本命かそうでないかはわからないので、事前確率を 0.5 と 0.5 と設定します。面積を用いて説明していきます。



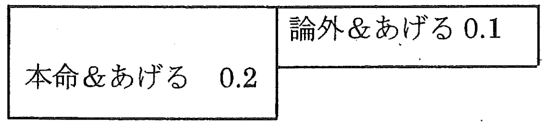
また、この本の中では、ある雑誌の働く女性へのアンケート調査にて、42.5%の確率で本命にチョコをあげ、論外（義理チョコ）にも 22%の確率であげるということが判明しました。
以下、条件付き確率を割り当てます。
ここで、条件付き確率が活躍します。
説明をしやすくするために、小数第 1 位までを考えます。

タイプ	チョコをあげる確率	チョコをあげない確率
本命	0.4	0.6
論外	0.2	0.8

面積を考えますので、かけ算によって、次の図のようになります。



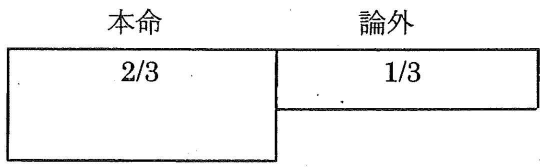
ここで、問題の設定として、チョコをもらったので、「あげない」という世界は消え去ります。



「あげない」可能性は考えなくてよいということ

よって、(左の長方形の面積)：（右の長方形の面積)

0.2　:　0.1
＝　2　:　1



この結果から、あなたがチョコをもらえた下で、あなたが彼女の「本命」である事後確率は、2/3＝約 66%となりました。

これが、ベイズ推定のプロセスです。
事後確率は、ベイズ逆確率ともいいます。

次ページ以降は、他の例を挙げます。

☆事前確率と事後確率についての例 1

3 つの袋があり、次のように赤い玉と白い玉が入っています。

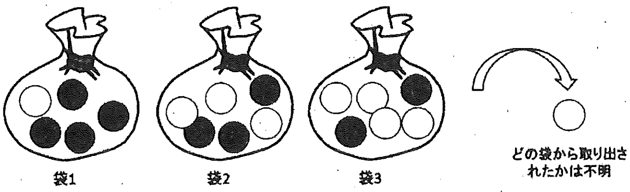


図 1

いずれかの袋から玉を 1 つ取り出したところ、白い玉でした。この玉が袋 2 から取り出された確率は、袋 2 から玉を取り出す事象を事象B_2、取り出した玉が白色である事象を事象 A とすると、ベイズの定理より次のように計算されます。

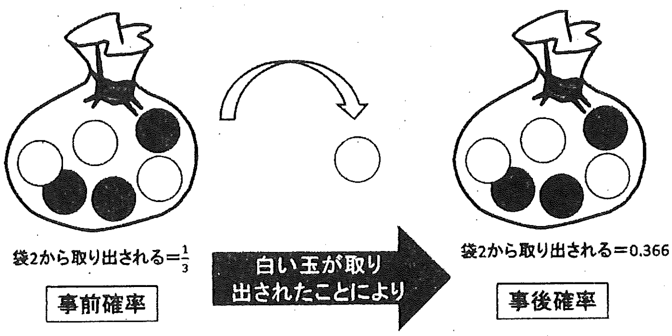
$$\begin{aligned} P(B_2|A) &= \frac{\frac{1}{3}\times\frac{1}{2}}{\frac{1}{3}\times\frac{1}{5}+\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times\frac{2}{3}} \\ &= 0.366 \end{aligned}$$

■事前確率

この計算では、取り出された玉の色が分からない、すなわち玉がまだ取り出されていないとき、袋 2 から玉を取り出す事象 B_2 の確率を $P(B_2)=\frac{1}{3}$ としています。これは、袋 1、2、3 の中からいずれか 1 つの袋を選ぶ確率は等しいということに基づきます。この確率は「事前確率」と呼ばれます。「事前」とは「事象 A が起こる前」ということを意味します。

■事後確率

いずれかの袋から玉を 1 つ取り出したところ白い玉であった $P(B_2|A)=0.366$ 場合に、袋 2 から玉が取り出された確率はとなりました。この確率は、「事象 A が起きた後」の、袋 2 から玉を取り出す事象B_2の確率であることから「事後確率」と呼ばれます。



この例は、高校生が、結果の確率が変化することを理解しやすいと思います。

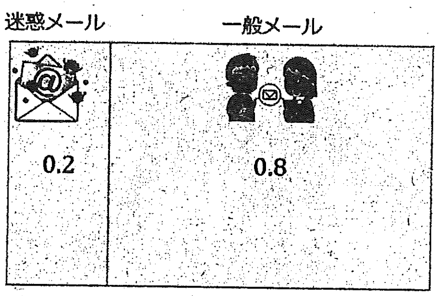
☆事前確率と事後確率の実生活への利用

迷惑メールを自動的に発見・分類する知恵

ベイズの定理(条件付き確率)が役に立っている代表例として、迷惑メールを自動的に発見・分類してくれるフィルタリング機能が挙げられます。

問題

過去の調査から、無作為に選んだメールの 20%が迷惑メール、80%が一般メールだと分かった。 調査によると、迷惑メールが『キャンペーン』という単語を含んでいる確率は 30%、一般メールが『キャンペーン』という単語を含んでいる確率は 4%である。 無作為に選んだメールが『キャンペーン』という単語を含んでいた場合、これが迷惑メールである確率は？



A：迷惑メールである

B：『キャンペーン』という単語を含んでいる

とおいて考えてみましょう。

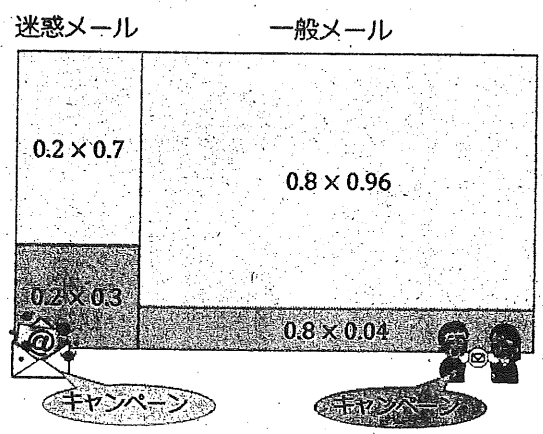
まず、過去の調査から「無作為に選んだメールが迷惑メールである確率」は P(A)＝0.2 だと分かっています。

次に、無作為に選んだメールが『キャンペーン』という単語を含んでいたという条件のもとで、それが迷惑メールである確率を求めます。

迷惑メールが『キャンペーン』という単語を含んでいる確率は 30%

一般メールが『キャンペーン』という単語を含んでいる確率は 4%

という情報を反映させてみましょう。



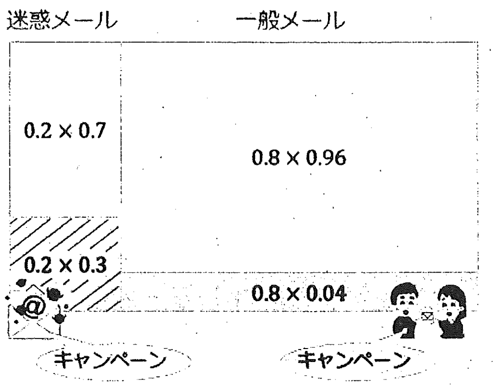
$$P(B) = 0.2 \times 0.3 + 0.8 \times 0.04 = 0.092$$

迷惑メール(0.2)という条件の下で『キャンペーン』という単語を含んでいる確率は 0.3 なので、「迷惑メールかつキャンペーンという単語を含んでいる確率」は 0.2×0.3

一般メール(0.8)という条件の下で『キャンペーン』という単語を含んでいる確率は 0.04 なので、「一般メールかつキャンペーンという単語を含んでいる確率」は 0.8×0.04 となります。

この図から、「無作為に選んだメールがキャンペーンという単語を含んでいる確率」は P(B)=0.2×0.3＋0.8×0.04＝0.092 だと分かります。(図の濃いグレー部分に相当)

あとは、下の図で、濃いグレー部分を分母、左下の斜線部を分子にとることで「無作為に選んだメールが『キャンペーン』という単語を含んでいたという条件のもとで、それが迷惑メールである確率」は P(A|B) ≒0.652 と求まります。



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.06}{0.092} \fallingdotseq 0.652$$

このように『キャンペーン』という単語を含んでいたことに着目することで、それが迷惑メールである可能性が 20%から約 65.2%、実に 3 倍にまで高まったことが示されました。

これを、ベイズ更新と言います。

3．最後に

条件付き確率が、人工知能などにも生かされていますので、生徒の学ぶ意欲の喚起につながる分野だと考え、現在、様々な知識を楽しみながら学んでいる最中です。

何かの参考にしてもらえたらありがたいです。

参考文献

[1] 小島寛之著　ダイヤモンド社
「完全独習　ベイズ統計学入門」

[2] 統計ＷＥＢ
事前確率と事後確率
https://bellcurve.jp/statistics/course/6446.html

[3] ベイズの定理とは何か。
条件付き確率からわかる判別の知恵
https://atarimae.biz/archives/15536