

現代数理統計学の基礎

～第1章 問6～

担当：いずびょん

2021年4月15日

概要

ベイズの定理を用いた基本的な問題です。後半には自分でもあんまり良く分かってない所を書いています。実際どうなんだろうと思っていることです。ご了承ください。

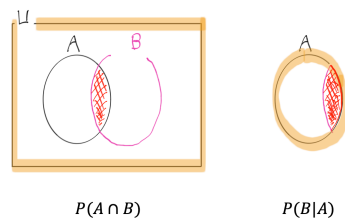
1 準備

1.1 条件付き確率

定義 1.3 (p.4) より、2つの事象 A と B があって $P(B) > 0$ (分母に持てきたいから) のとき、

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

を、 B を与えた時の A の条件付き確率という。



自分がつまづいたベン図。同じ所に注目してるけど、分母が違う。

1.2 ベイズの定理

命題 1.6 (p.5) より、 B_1, B_2, \dots を互いに背反な事象の列として、 $P(B_k) > 0, \bigcup_{k=1}^{\infty} B_k = \Omega$ を満たすとする。この時任意の事象 A に対して、 A を与えた時の B_j の条件付き確率 $P(B_j|A)$ は、

$$\begin{aligned}
P(B_j|A) &= \frac{P(A \cap B_j)}{P(A)} \\
&= \frac{P(A|B_j)P(B_j)}{P(\bigcup_{k=1}^{\infty}(A \cap B_k))} \\
&= \frac{P(A|B_j)P(B_j)}{\sum_{k=1}^{\infty} P(A \cap B_k)} (\because A \cap B_k, k = 1, \dots \text{は互いに背反}) \\
&= \frac{P(A|B_j)P(B_j)}{\sum_{k=1}^{\infty} P(A|B_k)P(B_k)}
\end{aligned}$$

これが Bayes の定理である。

2 回答

まず、「真に疾患がある」という事象を A 、「検査で陽性反応がでる」という事象を B とする。このとき、 A^c は「真に疾患がない」という事象を表し、 B^c は「検査で陰性反応がでる」という事象を表す。

与えられた問題の情報より、

$$P(B|A^c) = 0.2 \Rightarrow P(B^c|A^c) = 0.8$$

$$P(B^c|A) = 0.1 \Rightarrow P(B|A) = 0.9$$

$$P(A) = 0.1 \Rightarrow P(A^c) = 0.9$$

よって、求める確率 $P(A|B)$ は、Bayes の定理より、

$$\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(B|A)P(A)}{P((B \cap A) \cup (B \cap A^c))} \\
&= \frac{P(B|A)P(A)}{P(B \cap A) + P(B \cap A^c)} \\
&= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\
&= \frac{0.9 \cdot 0.1}{0.9 \cdot 0.1 + 0.2 \cdot 0.9} \\
&= \frac{1}{3}
\end{aligned}$$

3 合ってるか分からない変な考察

Bayes の定理の確率は、これまで習ってきた確率とは異なる。高校までは、全ての結果が等しく同じ確率で起こることから確率は定義された。大学では、繰り返し思考による頻度の極限として捉える頻度論を学んでいる。

今回の問題を例に整理する。まず、真に疾患がある確率は 0.1 であった。その後、検査で陽性反応が出た時、その確率は 0.33 に上がった。ここで、検査で陽性反応を受けるということが、疾患を患う確率を変化させうるであろうか？つまり、例えば検査薬に有害物質が入っていたのだろうか？

検査が有害ではないとき、検査で陽性反応を受けて変化したのは”現象”ではなく、ほかでもない我々の”認識”である。そしてこの”認識”は、客観的で科学的な結果から導かれた”認識”である。このような確率を、客観確率という（ちょっと違うかも？）。

ベイズ統計学とは、すでに起きた確実な事実であっても、それが確定的に知ることができない限り不確実とする考え方を体系的にまとめたものであり、客観確率はベイズ統計学の一つの捉え方である。反対に個人確率というもあり、個人確率は、勘や経験を用いて設定する。

今、最も知りたいのは「真に疾患がある」ことに関する確率であり、これを、検査の結果によって更新していく。このように、得られたデータによって、パラメータの確率（分布）をガンガン変化させるのがベイズ統計学の面白いところなのかな、と思う。反対に、「真に疾患がある」確率がしっかりと存在して、固定されているという考えのもとで、得られた検査の陽性反応の結果がそれに適合するかを検定したりするのが頻度論なのかなと思う。

この頻度論とベイズ統計学の2つの思想は、宗教論争になるくらい面白いらしい。あまり私には分からないが、頻度論者的には、今回の客観確率を用いるベイズは若干認めるものの、主観確率を用いたベイズはかなり受け入れられないらしい。ベイズは立場的には何も言わないで結果を淡々に積み上げていくイメージある（超主観）。

ちょっとごちゃごちゃしてしまいましたが、『モンテカルロ統計計算』を参考にしましたので興味がありましたらそちらを読んでください。