

ベイズ統計学概論

芝浦工業大学 数理科学研究会 BV19083 久保田静希

令和元年 11 月 4 日

1 研究背景

近年、コンピューターの発展により従来ではできなかった複雑な計算を誰もがたやすくできるようになった。これにより統計学は一部の人間のみが扱える高度な学問からパソコンをもっている者なら誰でも扱えるとても利便性の高いツールとなった。その中でも「ベイズの統計学」はその汎用性の広さから注目を浴びている。従来の統計学とは異なり、人間味のある計算を行うという点に惹か学習の着手に至った。

2 概要

ベイズの統計学の基盤となっている「ベイズの定理」から「ベイズ更新」等についての考察を行っていく。その中で、できるだけ頭の中でイメージが作りやすいように例を用いて解説していく。

3 ベイズの定理とは

ベイズの定理とは任意の事象 A, B に対し $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ が成り立つという定理である。ただしここで $P(A)$ とは事象 A が起こる確率を表し、 $P(B|A)$ は A が起きた時に B が起きる条件付き確率を表す。

4 ベイズの定理の解釈

ベイズの定理の事象 A を原因もしくは仮定 (Hypothesis)、事象 B をその A から得られたデータ (Data) と解釈する。こうして得られた式 $P(H|D) = \frac{P(D|H)P(H)}{P(D)}$ (D : データ, H : 原因) の左辺はデータ D が与えられたときその原因が H であった時の確率と解釈することができる。以下これを「ベイズの基本公式」と呼ぶ。

5 ベイズの定理の展開

ベイズの基本公式の $P(H|D)$ は D が起きた時その原因が H であった時の確立であるが、実際は考えられる原因は一つとは限らない。そこで考えられる原因が n 個あったとしそれぞれ H_1, H_2, \dots, H_n とする。この時 H_1 が原因であった時の確率は $P(H_1|D) = \frac{P(D|H_1)P(H_1)}{P(D)}$ と表せられる。またこれらの原因が全て排反であった時、 $P(D) = P(D \cap H_1) + P(D \cap H_2) + \dots + P(D \cap H_n)$ と表せる。ここで「確率の乗法定理」より $P(D) = P(D|H_1)P(H_1) + P(D|H_2)P(H_2) + \dots + P(D|H_n)P(H_n)$ と表すことができる。これを先ほどの式に代入し一般化すると

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{P(D|H_1)P(H_1) + P(D|H_2)P(H_2) + \dots + P(D|H_n)P(H_n)}$$

この式がベイズの統計学において鍵を握ってくる重要な式である。また、以下からベイズの定理とはこの式を指すものとする。

5.1 ベイズの定理の用語と意味

記号	名称	意味
$P(H_i)$	事前確率	原因 H_i が起こる確率
$P(D H_i)$	尤度	原因 H_i が起きた時 D が得られる確率
$P(H_i D)$	事後確率	D が得られたとき H_i が原因の確率

6 ベイズ更新

ベイズ更新とは従来の統計とは違い与えられた複数のデータの一つずつ処理していくベイズ統計学の醍醐味である。例として、あるデータ D_1, D_2 とその原因 H_1, H_2 が与えられたとする。このとき、二つのデータが与えられた原因が H_1 である確率を求めたいとする。このような場合まずは D_1 のみを用いてベイズの定理を利用し事後確率 $P(H_1|D_1)$ を求める。次に D_2 を用いてベイズの定理を使うのだがこの時、事前確率 $P(H_1)$ の代わりに先ほどの事後確率 $P(H_1|D_1)$ を用いて計算する。このような計算方法を行うことによって複数のデータに対し一つずつ処理を行うことができるのだ。これはデータ処理をするソフトウェアを作成するときに大変便利である。

6.1 理由不十分の原則

実際に日常でベイズ更新を用いようとするとき、一番最初の事前確率についての条件がないことが多々ある。しかし問題が不厳密であるからと言って問題を投げ出すことはできない。よってベイズの統計学では「特に条件がなければ確率は同等だろう」という常識を用いて計算を始める。ここが従来の統計学との大きな差である。

今後の課題

ベイズの統計学は事前確率に「理由不十分の原則」により自分たちの常識を用いてから計算を始めることが多い、それゆえに数学というには不厳密な点が多く残る。しかしながら、この不厳密さが人工知能や心理学などに応用されている。今後は不厳密さを更に利用して今までにない分野に応用することが求められている。

参考文献

- [1] 涌井良幸, 涌井貞美, 史上最強図解これならわかる! ベイズ統計学, ナツメ社, 2012
- [2] 涌井良幸, 涌井貞美, 史上最強図解これならわかる! 統計学, ナツメ社, 2010
- [3] 一石賢, まずはこの一冊から意味がわかるベイズ統計学, ベレ出版, 2016