魅力ある統計教育の教材開発2（ベイズ統計学の事始め）

林, 兵馬

(Citation)
研究紀要 : 神戸大学附属中等 論集,4:27-32

(Issue Date)
2020-03-31

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/81012812

(URL)
https://hdl.handle.net/20.500.14094/81012812

実践報告

魅力ある統計教育の教材開発 2(ベイズ統計学の事始め)

Developing an Attractive Teaching Material and Curriculum of Statistical Learning Ⅱ
(Starting Bayesian Statistics)

林　兵馬
HAYASHI Hyoma

本校数学科では，次期学習指導要領の改訂を見越し，本年度より統計教育に力を入れて行っている． 4年生(高校 1 年生)では，数学 I 「データの分析」数学 B「確率分布と統計的な推測」の一部を週 1 時間で学習しているが，本取組では，数学 A「場合の数・確率」の条件付き確率を用いて，データサイエンスにおいて重要性が高い「ベイズの定理」とその利用の基本を単元化し，実践を行った．その実践報告を行う．

From the academic year of 2019, the mathematics department of Kobe University Secondary School is focusing on statistics education based on the revision of the Course of Study that will be implemented in 2022. Currently, the fourth grade (senior high school first grade) students are supposed to study “data analysis” in the subject of Mathematics 1 and “probability distribution and statistical estimation” in Mathematics B once a week. This study designs a learning unit regarding “Bayesian theorem” and its basic use that has a higher level of importance in the field of data science by using conditional probability in the chapter “number of cases and probability “ in Mathematics A. This paper reports the lesson practice that was conducted based on this curriculum.

キーワード： データサイエンス，統計教育，条件付き確率，ベイズの定理
Key words:　data science, statistical education, conditional probability, bayesian theorem

I　はじめに

本校数学科では，現在統計教育を重点課題として推進している．

数学 A「場合の数・確率」の分野の「条件付き確率」の内容で，一般に原因の確率として知られている内容を「ベイズの定理」として単元構成を行い実施した．

II　ベイズの定理

1.ベイズの定理とは

「観測事象（観測された事実）から、推定したい事柄（それの起因である原因事象）を、確率的な意味で推論」するために，必要な定理がベイズの定理である．

マイクロソフトのビルゲイツは自社が競争上優位にあるのはベイズ統計によると宣言したり、グーグルでは検索エンジンの自動翻訳システムでベイズ統計の技術を活かしていることが知られている．

「観測事象（観測された事実）から、推定したい事柄（それの起因である原因事象）を、確率的な意味で推論」するために，必要な定理がベイズの定理である．

2.ベイズの定理の授業での導出

まず，本取り組みで最重要の定理である「ベイズの定理」を授業にどのように導出したか説明を行う．

全事象 U の中の２つの事象 A と B について，A が起こったとわかったとして，このとき B が起

こる確率を，A が起こったときの B の条件付き確率といい，$P(B|A)$で表す．（ただし，高等学校数学 A の教科書では，$P_A(B)$で表現するが，本紀要では，より一般的な表記である $P(B|A)$で表現する．）

条件付き確率 $P(B|A)$は，全事象を A としたときの B が起こる確率であるので，

$$P(B|A) = \frac{n(A \wedge B)}{n(A)} = \frac{P(A \wedge B)}{P(A)} \cdots ①$$

となる．

ここで，①より

$$P(A|B) = \frac{P(B \wedge A)}{P(B)} \cdots ②$$

が得られる．また，

$$P(A \wedge B) = P(B \wedge A)$$

と，①，②より次の定理が得られる．

ベイズの定理

P(B) > 0のとき

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)} \cdots ③$$

また，$P(B) = P(A \wedge B) + P(\bar{A} \wedge B)$

$$= P(A) \times P(B|A) + P(\bar{A}) \times P(B|\bar{A})$$

であるので，③の別表記が得られる．

ベイズの定理

P(B) > 0のとき

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(A) \times P(B|A) + P(\bar{A}) \times P(B|\bar{A})} \cdots ④$$

一般的に，A に観測事象，B にそれが起因である原因の事象になるので，$P(B)$を直接求めることができない．

しかし，Bの起こりうることを場合分けし求めることにより，計算ができる．

Ⅲ　授業実践

教科書での基本的な条件付き確率の指導が終わった後，ベイズの定理の証明とその利用に入った．なお，取り扱う問題は，「見えないものをさぐる―それがベイズ: ツールによる実践ベイズ統計(オーム社)」を参考にした．

まず，問題$\boxed{1}$に取り組んだ．

問題$\boxed{1}$　(1)

ある病気を発見する検査法は，その病気の患者が受けると 99%の人が陽性反応を示します．

しかし，その病気でない健康な人でも 3%の人が陽性反応を示し，誤った診断を下してしまいます．

日本国内で，その病気の患者の割合は 0.1%であることが分かっています．

さて，日本に住む E さんがこの検査を受けたら陽性反応がでました．E さんがその病気である確率は何%か？

まず，ベイズの定理を用いずに求めるように授業を行った．

［解答］ベイズの定理を用いない方法

まず，日本国内の人口を 100,000 人とする．

すると，以下のように場合分けができる．

・病気の人　　：100,000×0.001=100(人)

・病気でない人：100,000×0.999=99,900(人)

①　病気の人が検査を受ける場合

陽性：100×0.99=99(人)

陰性：100×0.01=1(人)

②　病気でない人が検査を受ける場合

陽性：99,900×0.03=2,997(人)

陰性：99,900×0.97=96,903(人)

求める確率は，

$$\frac{\text{陽性の中で病気の人}}{\text{陽性の人}}$$

であり，陽性の中で病気の人は，病気の人の中で陽性の人であるので，

$$\frac{99}{2{,}997+99} = 0.032$$

よって，3.2%.(終)

小集団活動で求めさせたが「病気の人の中で陽性」「病気の人の中で陰性」「病気でない人の中で陽性」「病気でない人の中で陰性」の数を求めたのちに求める確率を計算するように指示をしたところ，ほとんどの小集団が求値できていた．この後にベイズの定理の証明を行い，その中でも重要になる，

なお，最初に日本に 100,000 人いるとして確率を求値したが，即ベイズの定理を用いるよりも下記の点でよいと考えられる．

・１回目の検査で陽性の人が病気である確率が

3.2%と，あらかじめ予想した値よりも低い理由が

$$\frac{99}{2{,}997+99}$$

の 2,997 が大きく影響していることが一目瞭然であり，また病気でない人のうち陰性判定する確率が 3%であることに起因していると多くの生徒が実感できたと考えてられる．

・$\frac{99}{2997+99}$より，

$$\frac{\text{陽性の人のうち，病気の人}}{\text{陽性の人}}$$
$$=\frac{\text{陽性の人のうち病気}}{\text{陽性の人のうち病気の人}+\text{陽性の人のうち病気でない人}}$$

であることが強く意識づける指導をし，

$$\frac{P(A)\times P(B|A)}{P(A)\times P(B|A)+P(\bar{A})\times P(B|\bar{A})}$$

が，数式として何を主張しているのかを理解させるように心がけた．

また，

$$P(A\wedge B)=P(B\wedge A)$$

の関係を意識させるために～～～を全体共有では強調した．問題$\boxed{1}$を，ベイズの定理を用いて解く前に，次の問題を練習として挟んだ．

問題$\boxed{2}$

ある学年に生徒が 100 人います．学年の生徒の中でカメラと時計の両方を持っている確率を求めたい．

この学年の中の御影くん，住吉さん，明石さんの 3 人に, 100 人の中でカメラと時計を所有している人をそれぞれ調べてもらいました．

この 3 人の中の情報で，学年の生徒の中でカメラと時計の両方を所持している確率を求められるものはどれか．

・御影くん

学年 100 人の中で,

カメラを持っている人は 50 人

時計を持っている人は 60 人

カメラと時計の両方を所持しているのは 20 人

・住吉さん

学年 100 人の中で,

時計を持っている人は 60 人

時計を持っている人のうち，カメラを持っている人は 20 人(カメラを持っている人は不明)

・明石さん

学年 100 人の中で,

カメラを持っている人は 50 人

カメラを持っている人のうち，時計を持っている人は 20 人(時計を持っている人は不明)

[解答]

全事象：学年全員

事象 A：時計を持っている人

事象 B：カメラを持っている人

・御影くん

$$P(A\wedge B)=\frac{20}{100}$$

・住吉さん

$$P(A\wedge B)=P(A)\times P(B|A)=\frac{60}{100}\times\frac{20}{60}=\frac{20}{100}$$

・明石さん

$$P(B\wedge A)=P(B)\times P(A|B)=\frac{50}{100}\times\frac{20}{50}=\frac{20}{100}$$

よって, 3 人全員の調査から求める確率を求めることができる．(終)

問題$\boxed{2}$の難易度は平易であるが，$P(A\wedge B)$を 2 種類の条件付き確率

$$P(A\wedge B)=P(A)\times P(B|A)$$
$$P(B\wedge A)=P(B)\times P(A|B)$$

を用いて表すことを具体例で実感させることを目的としている．その後，上記Ⅱベイズの定理の 2.ベイズの定理の授業での導出で記述した方法で，全体共有を行った．

その後，ベイズの定理を導出したので，問題$\boxed{1}$(1)をベイズの定理を用いて全体で共有した．

[解答]ベイズの定理を用いる方法

事象 A：E さんが病気である

事象 B：陽性である

求める確率$P(A|B)$は

$$P(A|B)=\frac{P(A)\times P(B|A)}{P(A)\times P(B|A)+P(\bar{A})\times P(B|\bar{A})}$$

であるので，これに代入する．

P(A) : (検査を受ける前の)E さんが病気である確率

→E さんは日本に住んでいるので，0.001

$P(\bar{A})$：上記より 0.999

$P(B|A)$: (E さんが)病気であるときに，陽性である

確率→0.99

$P(B|\bar{A})$: (E さんが)病気でないときに，陽性である確率→0.03

よって，

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(A) \times P(B|A) + P(\bar{A}) \times P(B|\bar{A})}$$
$$= \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.03} = 0.032$$

よって，3.2%.(終)

上記にあるように，問題$\boxed{1}$(1)のベイズの定理を用いる解法と，用いない解法の間に問題$\boxed{2}$を挟み，ベイズの定理の理解を深めるように心掛けた.

問題$\boxed{1}$(2)

(1)の検査で陽性反応を示したので，心配になった E さんは，最初の検査を受けた病院で紹介状を書いてもらい，別の病院で検査を受けることにしました.

この病院の検査は以下の通り.

・この病院の患者が検査を受けると 98%の人が陽性反応を示し，その病気でない人でも 2%の人が陽性を示す.

・この病院はその病気を専門的に取り扱っており，紹介状がなければ来院出来ない．紹介状を持って来院する人はその病気の疑いが高く，この検査でおおかた病気かどうか判断できる.

E さんがこの病院で検査を受けると再度陽性反応を示した．このとき，E さんが病気である確率は何%か？

問題$\boxed{1}$(2)は，ベイズの定理による確率更新(ベイズ更新)を理解させる狙いがある．問題$\boxed{1}$(1)で求めた事後確率 3.2%を(2)では事前確率にするところがポイントである．求めた事後確率を次の事前確率に用いることを理解させたい.

[解答]

事象 A : E さんが病気である

事象 B : 陽性である

$P(A)$: (2 回目の検査を受ける前の)E さんが病気である確率

→E さんは 1 回目の検査で陽性反応を示し，病気である確率が，0.032 と求まったので，これを事前確率に用いる.

$P(\bar{A})$: 上記より 0.968

$P(B|A)$: (E さんが)病気であるときに，陽性である確率→0.98

$P(B|\bar{A})$: (E さんが)病気でないときに，陽性である確率→0.02

よって，

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(A) \times P(B|A) + P(\bar{A}) \times P(B|\bar{A})}$$
$$= \frac{0.032 \times 0.98}{0.032 \times 0.98 + 0.968 \times 0.02} = 0.618$$

よって，61.8%.(終)

2 回目の検査は，紹介状を持って来院しなければ検査が受けられない状況で，また 2 回連続で陽性反応を示したことにより,E さんが病気である確率が急に上がった様子を共有した.

また，この時点でベイズの定理と確率更新について以下の注意を全体に行った.

注意 1

ベイズの定理は，

P(B) > 0のとき

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

であるが，

$P(B|A)$は，A が起こる前提で B が起こる確率

$P(A|B)$は，B が起こる前提で A が起こる確率

であり，片方が求まっていると，その逆向きが求まることを示している.

注意 2

$P(A)$を**事前確率**，$P(A|B)$を**事後確率**とよぶ.

注意 3

観測のたびに，0.1%から 3.2%へ，3.2%から 61.8%へと確率が更新されていっている．このように，観測を行いながら事後確率を更新していくことを**ベイズ更新**という.

問題$\boxed{3}$

村はずれの羊使いの少年が，退屈だったのでオオカミがいないのに，「オオカミがきたぞ！」とウソをついて騒ぎを起こします.

村人たちは，これは大変とオオカミ退治に向かいますが，いくら探しても肝心のオオカミがいま

せん．これをおもしろがった少年は，このウソを繰り返します.

ところがある日，本当にオオカミがやってきました.

少年は「オオカミがきたぞ！」と伝えましたが，村人はもうその少年の言葉を信用せずオオカミ退治に行かなかったため，村の羊はすべてオオカミに食べられてしまいました.

(1)　村人，ここでは村の代表者である村長は，何回目のウソでこの少年を信用しなくなったのか？

ここまでベイズの定理を扱ってきたが，少年が「オオカミがきたぞ！」といったが，オオカミがこなかった理由に「少年がウソをついたことにより村人はオオカミと出会えなかった」場合以外に，「少年は本当のことを言ったが，(なんらかの理由で)村人はオオカミと出会えなかった」ことがある.特に後者の状況は，丁寧に全体共有する必要があるように感じた.

また，後述する各種条件の正確な確率は求めることができないので，**主観確率**とよばれる主観的な確率を設定し，ベイズの定理で確率を更新していく．また，少年は「オオカミがきたぞ！」とウソをつき続けるものである(途中で本当のことを言わない)ことも全体で十分に確認しておく必要がある.

なお，生徒に「少年がウソつきだ」と村長が思う確率がいくつを超えると，ウソつきと判定するか？と問いかけてみると，0.8 や 0.9 といった答えが返ってきた. 今回は 0.9 を超えるとウソつきと判定することにする.

[解答]

事象 A：少年はウソつき

事象 B：オオカミはやってこない

あらかじめ必要な事前確率と条件付き確率を設定する.

・$P(A)$少年がウソをつく以前，村長が「少年はウソつきだ」と思う事前確率(主観)→0.1

$P(B|A)$:少年がウソつきであったとき，オオカミがこない(村人はオオカミと出会えない)確率→0.95

$P(B|\bar{A})$:少年が本当のことを言ったとき，オオカミがこない(村人はオオカミと出会えない)確率→0.3

この設定で，オオカミと出会えないとき，少年がうそつきである(うそつきであることに起因する)確率$P(A|B)$は，

$$P(A|B)=\frac{P(A)\times P(B|A)}{P(A)\times P(B|A)+P(\bar{A})\times P(B|\bar{A})}$$

から求める.

また，2 回目以降の事前確率は，その 1 回前の事後確率を用いる.

すると，

この手順で帰納的に事後確率を求める.

1 回目:26.0%

2 回目:52.7%

3 回目:77.9%

4 回目:91.7%

5 回目:97.2%

6 回目:99.1%

7 回目:99.7%

よって，0.9(=90%)を超えるとウソつきと判定することにするならば, 4 回ウソをついたときである.

(終)

(2)　もし，初回が「本当」の場合(すなわち，少年は本当にオオカミを発見し，実際に村人がオオカミと出会えた場合)，その後少年はウソをつき続けてオオカミと出会えなかった場合，何回目のウソで村長はこの少年を信用しなくなったのか？

この問題のポイントは，以下の 2 点である.

・1 回目の本当が，後のウソにどのように影響するのか.

・1 回目の本当であった現象を

事象 A：少年はウソつき

事象 B：オオカミはやってこない

を用いてどのように記述するのか.

[解答]

最初に本当のことをいい，実際にオオカミと出会えた

$$P(A|\bar{B})=\frac{P(A)\times P(\bar{B}|A)}{P(A)\times P(\bar{B}|A)+P(\bar{A})\times P(\bar{B}|\bar{A})}$$
$$=0.0078=0.78\%$$

これを用いて，(1)と同様に計算すると，

1回目:2.5%
2回目:7.4%
3回目:20.1%
4回目:44.9%
5回目:71.6%
6回目:88.9%
7回目:96.2%　　　　　となる.

よって，0.9(=90%)を超えるとウソつきと判定することにするならば, 7回ウソをついたときである. (終)

(3)　一度も少年は本当のことを言わなかったとき，村長は，少年の何回目のウソでオオカミ退治に「行かない」選択をすべきか．損失の期待値を計算することで判断しなさい.

本来は，条件付き確率で処理をする方がより正確ではあるが，簡便的に損失の期待値を以下のように計算した.

①　オオカミ退治に「行かない」選択をした場合

・オオカミが来た場合の損失：10,000,000 円(羊がすべてやられる)

・オオカミが来なかった場合の損失：0円

よって，この場合の損害の期待値は,

$$10{,}000{,}000 \times \frac{9}{10} + 0 \times \frac{1}{10} = 9{,}000{,}000$$

②　オオカミ退治に「行く」選択をした場合

・オオカミが来た場合の損失 : 200,000円(見回りの費用)＋500,000円(オオカミとの戦闘)=700,000円

よって，この場合の損害の期待値は,

$$700{,}000 \times \frac{9}{10} + 200{,}000 \times \frac{1}{10} = 650{,}000$$

したがって，この場合は，オオカミ退治に「行く」選択をすべきである.

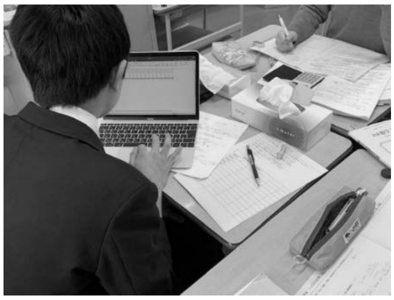
同様に

	損害の期待値	
	行かない場合	行く場合
1回目のウソの後	7,398,000円	569,900円
2回目のウソの後	4,730,000円	436,500円
3回目のウソの後	2,209,000円	310,450円
4回目のウソの後	822,000円	241,100円
5回目のウソの後	275,000円	213,750円
6回目のウソの後	89,000円	204,450円
7回目のウソの後	29,000円	201,450円

6回目のウソの後では, 損害の期待値がオオカミ退治に「行かない」場合よりも「行く」場合の方が高いことが分かる.

したがって, 7回目以降の「オオカミがきたぞ！」という連絡には応じなくてもよい. (終)

[授業の様子]



[(自発的に) PCで損害の期待値を求めている]

Ⅳ　評価

秋学期期末考査にて，ベイズの定理とその利用に関して出題を行い，評価と検証を行う予定であったが，新型コロナウイルスにより，考査を実施することは出来なかった．次年度の課題とする.

Ⅴ　謝　辞

研究紀要発刊にあたり, (株)日立システムズ板井光輝氏，本校数学科の先生をはじめ多くの方にご指導を頂いた．厚くお礼申し上げる.

文　献

藤田一弥 2015『見えないものをさぐる―それがベイズ: ツールによる実践ベイズ統計』(オーム社)