# 意思決定におけるベイズ理論の有用性
## ―高等学校における確率・統計教育への導入の検討―

弘前大学大学院　教育学研究科
教科教育専攻　数学教育専修
14GP206　葛西　晃

# 目 次

# 序章　概要

## 第 1 節　背景・目的

日本学術会議によって平成 27 年 12 月 17 日に報告された大学教育の分野別質保証のための教育課程編成上の参照基準 [統計学分野] において、統計学分野を学ぶすべての学生が身に付けるべき基本的な素養を次のように述べている。

> 統計学は、歴史的には、不確実性の下において賢明な意思決定を行うための科学として発展してきた。統計学を学ぶ本質的意義の一つは、付録に見るように、自然界には本質的に不確実性が存在するため、自然現象の理解のためには、確率論や統計学が重要な役割を果たしていることを認識し、自然を理解するための適切な方法を習得することである。また、人間社会の諸問題においても、様々な不確実性が存在し、それに賢明に対処するためには、統計学の考え方や課題解決の手順が重要な役割を果たしていることを認識し、それに対処するための方法を習得することも、統計学を学ぶ本質的意義と考えられる。
>
> 次に、課題解決型の思考力にとって必要なことは、物事を分析・整理し、帰納的・演繹的な推論により要素間の再結合や別の知識との融合・再構築により新たな考えを生み出すことで、物事の間の既存の論理的な鎖を越えて飛躍し、融合・再構成された概念の下に整合性の取れた新しい論理系を作り上げることと言うことができる。この過程は、上述した統計学を用いて現実問題に対処する過程と同様のものである。したがって、統計学を学ぶことにより、データや資料の情報に基づいて論理的な分析・課題解決を行うことのできる課題解決型の思考力を獲得することが可能となる。これもまた、統計学を学ぶ本質的意義と考えられる。
>
> さらに、また、より良き市民生活をおくるためには、データを基に、様々な場面で遭遇する不確実性に適切に対処するための、リスクを最小にする判断を行うことが要求される。このためには、一人一人が統計的素養（統計リテラシー）を持っていることが不可欠であり、統計学に固有の知的訓練は、これに大きく寄与することができる。これもまた、統計学を学ぶ意義の一つである。(日本学術会議,2015,p.8)

このことから、統計学分野を学ぶすべての学生が身に付けるべき基本的な素養は以下の 3 点であることがわかる。

(I) 自然界、人間社会の諸問題に存在する不確実性に対処するために、統計学の考え方が重要な役割を果たすことを認識し、その方法を習得すること。

(II) データや資料の情報に基づいて論理的な分析・問題解決を行うことのできる問題解決型の思考力を獲得すること。

(III) データを基に、様々な場面でリスクを最小にする判断を行うための統計的素養を養うこと。

また、参照基準によれば、統計学は、データをもとに現象を記述し、現象のモデルを構築し知識を獲得するための方法論であるが、この方法により不確実性を伴う現象の予測や制御、またリスクを考慮した際の合理的な意思決定が可能である、とあるように、統計学を学ぶことによって、合理的な意思決定ができるようになることを期待できる。また、その意思決定はある専門的なものに限るのではなく、ほとんどの学問分野の研究活動や実社会など、とても汎用的な方法であるため、従って、統計学は意思決定を考える際には誰しもが必要とする学問の1つと言える。そして、先程述べた3点は、より賢明な意思決定を行うためには必要不可欠なものであると思われる。

ところで、文部科学省が平成28年12月22日に公表した学校基本調査によると、高等学校（全日制・定時制）卒業者の内、大学・短期大学進学率は平成28年度で56.8%(文部科学省,2016,p.4)であり、高等学校卒業後、大学・短期大学へ進学しない人も多くいることがわかる。また、文部科学省が平成28年4月13日に公表した高等学校の教育課程に関する基本資料によると、高等学校等への進学率は平成27年度で96.6%(文部科学省,2016,p.6)である。このように大学・短期大学進学率と高等学校進学率を比べても、明らかに高等学校進学率が高いことがわかる。

現在、高等学校で実施している確率・統計教育は、頻度主義を基にした確率・統計学を展開している。しかし、人間社会の諸問題における意思決定に頻度主義では馴染まない問題が多々存在する。従って、社会生活を送る多くの人が最後に統計学を学ぶ高等学校において、頻度主義によらない意思決定における統計的素養をきちんと養う必要があると考える。

第1章で詳しく述べるが、ベイズ理論の歴史によれば、意思決定におけるベイズ理論の有用性を推察することができる。

これらを踏まえ、本研究の目的は、

- ベイズ理論が、意思決定において、頻度主義を基にした確率・統計学に比べ、適応できる場合が多い理論であることを示すこと

- 高等学校の確率・統計教育への導入が可能であることを示すこと

とする。

## 第 2 節　構成

第 1 章第 1 節では、ベイズ理論の歴史について述べる。

第 1 章第 2 節では、第 1 章第 1 節で得られたベイズ理論の歴史からどのようなことが言えるのか考察する。

第 2 章第 1 節では、ベイズ理論の基となるベイズの定理を紹介し、この定理の証明をする。

第 2 章第 2 節では、ベイズの定理を用いることで得られる事後確率をもう一度、事前確率に置き直し、データから新たな事後確率を導くベイズ更新について説明し、一般化する。

第 2 章第 3 節では、事後確率と損失関数で表すことができるリスクを定義する。

第 2 章第 4 節では、意思決定にベイズ理論を利用した具体例を 2 つ紹介する。

第 3 章第 1 節では、決定理論を紹介する。

第 3 章第 2 節では、決定方式を選ぶ際の順序付けと許容性について整理する。

第 3 章第 3 節では、意思決定方式であるミニマックス決定方式とベイズ決定方式の 2 つを定義し、ベイズ決定方式から得られる事後期待損失最小化の定理を証明する。

終章第 1 節では、第 1 章から第 3 章の結果を受け、総括を述べる。

終章第 2 節では、総括から判断できる今後の課題を述べる。

本研究の内容については以上であるが、ベイズ理論を基にした統計学（以降、ベイズ統計学と呼ぶ）が頻度主義の統計学と同様に扱える保証するために付章を加える。

付章第 1 節では、ベイズの定理を確率の形から統計学の形へと発展させる。

付章第 2 節では、情報理論におけるカルバック・ライブラーの情報量を定義し、性質について証明する。

付章第 3 節では、カルバック・ライブラーの情報量を用いることで、ベイズ統計学が頻度主義を基にした統計学と同様に扱えることを明らかにする。

# 第１章　ベイズ理論の歴史と考察

本章の構成

第１章第１節では、ベイズ理論の歴史について述べる。なお、この節は、著者シャロン・バーチェ・マグレイン、訳富永星、『異端のベイズ統計学』（2013）に基づいて述べる。

第１章第２節では、第１章第１節で得られたベイズ理論の歴史からどのようなことが言えるのか考察する。

## 第１節　ベイズ理論の歴史について

〇 1740 年代～1764 年

1740 年代、トーマス・ベイズによってベイズの定理の基となる考え方が発見される。しかし、この発見は世間に知られることなく、ベイズは 1761 年にこの世を去ってしまう。ベイズの死の 3 年後、友人であるリチャード・プライスがその考え方が載っている論文を論文の山の中から発見し、遺作として出版した。その出版された論文『An Essay Toward Solving a Problem in the Doctrine of Chances（訳　確率の考え方における、ある問題を解くための試論）』には、原因の確率に関する考え方が書かれており、その形は現在、ベイズの定理と呼ばれている形のものではなく、あくまでベイズの定理の基となる考え方が書かれていただけであった。

〇 1773 年～1827 年

ピエール・シモン・ラプラスは、太陽系の不確かさといった天文学の問題を解決するために確率が役立つと考えた。どのように役立つのかというと、天体の動きは複雑で、正確な解は得られそうにないが、確率を使うことで、天文学の誤差だらけの観測データの中で、どのデータの方が正しそうかといった程度はわかる、と考えた（実際は確率を知識を使って説いたわけではなかった）。ラプラスは、天体の運動のうちに観察される変則性の原因を発見する方法や、観測の誤差のうちに隠されている小さな不変性の原因を識別する、推定する方法を考え始めた。そして、すでにわかっている出来事に基づいて、そのもっともありそうな原因へとさかのぼるための数学的な一般理論を作るべく、手探りで前進した。ラプラスは、このような着想を「原因の確率」と呼び、確率の研究を進めた。1774 年、この着想を式にせず、直観的に原理として、「（ある出来事があったときに、その）原因の確率は、その出来事が起きる確率に比例する」と発表し、そして、1810 年から 1814 年にかけて、今日、ベイズの定理と呼ばれている定理を定式化することに成功した。さらに、ラプラスは、この定理を作る中で、自分の直観（主観確率）を客観的なデータで更新する手法も用いていた。具体的には、男女出生比の調査において、はじめは出生比を等確率で考え、それを客観的なデータで更新していき、男女出生比を求めていった。

○ 1827 年〜1930 年代

ラプラスの死後、ベイズ理論への激しい批判が始まる。このような変化が起きたのは、(1) 十分な証拠がないにもかかわらず、ラプラスの名声をなじる声が広がってしまったこと、(2) 客観説が美徳とされ、主観確率といったベイズ理論の主観説が攻撃の的となったこと、の 2 点が原因である。（ここで、述べている客観説とは、確率が相対頻度的定義、つまり、$A$ という事象を、$A$ の起った回数の相対頻度の極限 $P(A) = \lim(\frac{n_A}{n})$ で求められることが前提となっていることである。従って、主観説とは、この前提を必要としないことである。）これをきっかけにして、頻度主義が台頭していく。1900 年頃からは、頻度主義が中心となり推測統計学が一気に発展していった。カール・ピアソン、イェジー・ネイマンなどの立役者がいる中で、特に、ロナルド・フィッシャーの功績が頻度主義の地位を高めていった。

ローサムステッド農業試験場で化学肥料を分析する仕事に就いていたフィッシャーは、農業に関わる大量のデータを分析していた。このとき、新たなタイプの実際的な問題を突き付けられ、実験を設計するより良い方法を探し始めた。そして、長い時間をかけて、有意性検定、分散分析などを作り出した。1925 年に『Statistical Methods for Research Workers（訳　研究者のための統計学的方法）』を発表し、頻度主義が事実上の標準的な統計的手法となった。

○ 1939 年〜1940 年代後半

1939 年頃になると、頻度主義の影響はさらに大きくなり、ベイズ理論は事実上タブーとされていた。しかし、第二次世界大戦が始まったことにより、戦時下の指導者たちは、完全な情報が得られるのを待たずに、人の生死に関わる意思決定を素早く最良の形で行う方法を考える必要があった。そんな中、ドイツ軍のエニグマの暗号を解読するために、イギリス政府は言語学者以外に数学者を数人、集めていた。その中にいた、数学者であるアラン・マシスン・チューリングは暗号解読にベイズ理論を用いていた。また、暗号解読にベイズ理論を用いていたのは、チューリングだけでなく、アメリカの数学者であるアンドリュー・マッティ・グリーソンも利用していた。第二次世界大戦において、ベイズ理論は大きな貢献をした。それにもかかわらず、機密扱いとなり、この功績が世間に知られることはなかった（世間に知れたのは 1974 年以降であった）。

また、1940 年代に、エイブラハム・ワルドは、ゲーム理論の観点を統計理論に導入することを考え、統計的決定関数の理論を作り上げた。ワルド自身はベイス理論派ではなかったが、統計的決定関数の理論を構築する過程で、ベイズ理論を用いるようになり、ベイズ理論を用いた方法がどのように優れているのかを数学的に示した（完備類定理）。

○ 1950 年〜1960 年代

第二次世界大戦後、統計学ブームによってベイズ理論の地位も徐々に押上られて

いった。そんな中、ベイズ理論の基礎を構築していったのが、ジャック・グッド、レオナード・ジミー・サヴェジ、デニス・V・リンドレーの3人であった。

グッドはエニグマの暗号を解読したチューリングの助手であり、エニグマの暗号解読に用いたベイズ理論に関する論文を書いた。1950年に、『Probability and the Weighing of Evidence（訳　確率と証拠の重みづけ）』を発表するも、自分の考えを教えたり、説明したりするのが不慣れである性格と、また、第二次世界大戦中のことについては機密保持が義務づけられていたため、注目されることがなかった。そのため、他の2人がベイズ理論派の知的リーダーとなった。

サヴェジは経済学者ミルトン・フリードマンとの仕事をきっかけに統計学を研究するようになる。統計学を研究する中で、1954年に、『The Foundations of Statistics（訳　統計学の基礎）』を出版した。この本によって、サヴェジはベイズ理論の基礎となる主観確率（サヴェジは個人的確率と呼んでいた）の意味付けに大きな貢献をしており、ベイズ理論の知的リーダーの1人となった。

リンドレーはイギリスの確率論の中心地であり、フィッシャーやチューリング、グッドが働いたり、学んだりした場所であるケンブリッジ大学で統計学の研究していた。1954年、シカゴ大学を訪れた際に、サヴェジと出会ったことで、主観確率、ベイズ理論の正当化を求めるようになった。そして、ベイズ理論を基にした統計学を広めるための運動の指導者、推進者となった。

この3人（サヴェジとリンドレーの2人の功績が大きいが）よって、ベイズ理論は論理的な基礎を築くことができたが、このベイズ理論を意思決定に使おうとする者が現れた。それが、ロバート・オシャー・シュレイファーとハワード・ライファであった。

ハーバード大学ビジネススクールで働いていたシュレイファーは「統計的品質管理」の講座を割り当てられた。統計について何も知らなかったシュレイファーは、フィッシャー、ネイマン、ピアソンといった頻度主義の理論家たちの著作を読んで知識を仕込んだ。その時、新製品の投入や価格の変更などの問題を解決することに、頻度主義がまるで役に立たなかった。その理由は、頻度主義の方法が、反復可能な現象の場合には有効な方法であるからと理解した。シュレイファーは、データが全くない中で、企業の幹部たちはどうやって意思決定をしているのだろうかと考えた。そして、事前情報をサンプルデータに基づいて更新していく、ベイズ理論が使えると考えた。

コロンビア大学で働いていたライファは、統計学部において、ネイマン・ピアソンの理論や仮説の検定、信頼区間などをテーマに教えていた。その中で、統計学をデータ分析だけでなく、意思決定にも使えないものかと考えるようになった。意思決定を考える中で、ライファが関心を持ったのは、1回しかない具体的な決定、例えば、製品をどのくらい備蓄するかとか、どれくらいの価格をつけるかといった素

早い判断を必要とする決定であった。さらに、「それまでの意見に基づいて、また、具体的なサンプル事象に照らして、$p$ 値（$p$ 値とは、帰無仮説の下で、実際にデータから計算された統計量よりも極端な（仮説に反する）統計量が観測される確率のことである）が 0.25 より大きい確率は 0.92 だと考えられる」ということを事業主は言えるようになりたいと思っていると、ライファは考えていた。従って、仮説を棄却したり、採択するだけでは不十分ではないかと考えていた。そして、ライファはベイズ理論派へとなり、意思決定問題にベイズ理論を用いた。

〇 1970 年代〜2000 年代

1970 年代に入ると、アメリカでは、1971 年にサヴェジがこの世を去り、イギリスでは、1977 年にリンドレーが管理職を降りて、単独で研究を始めるなど、指導者が不在となり、幾人もの転職者が続いたことが原因となり、ベイズ理論は停滞期に入る。しかし、1980 年年代前半から、コンピュータのハード面、ソフト面の発達により、ベイズ理論を用いた統計学が注目されるようになる。計算機器が発達していなかった頃は、ベイズ理論が、主観確率という思想が問題視されただけでなく、ごく簡単なモデルの場合を除き、肝心な事後分布を求めることが難しく、実用的ではないという問題があったが、この問題はかなり解消した。特に、マルコフ連鎖モンテカルロ法（ＭＣＭＣ）と呼ばれる手法がある。ＭＣＭＣの起源は 1950 年代に遡るが、統計物理学などの他分野で開発されたものであり、これがベイズ理論にも大いに活用できることが 1989 年に知られた。そして、1991 年に、ＢＵＧＳという無料のプログラムが開発されたのを皮切りに、ベイズ理論を用いた統計学用のＭＣＭＣのソフト面は整備された。こうして、計算面におけるベイズ統計学の難点は 20 世紀の終わりごろに解消され、その後、様々な分野でベイズ理論が活躍している。

## 第 2 節　歴史についての考察

歴史からベイズ理論はどの年代においても、主観確率を用いていることで頻度主義と考えられる人々から批判され、衰退していく過去があったと読み取ることができる。しかし、1940 年代から、ワルドの統計的決定関数の理論によって、統計学を、自然界の法則を理解することだけでなく、人間の行動決定にも用いるようになったと判断できる。そして、その場合、ベイズ理論を用いたことがわかる。また、ワルド、シュレイファー、ライファという意思決定にベイズ理論を利用した人たちは、全員、初めからベイズ理論を支持していたのではなく、頻度主義を基にした統計学を学んでいたり、教えていたにもかかわらず、結果として、意思決定に頻度主義の統計学ではなくベイズ理論を重視していた。従って、ベイズ理論は頻度主義に比べ、意思決定の分野においては有用であると結論できる。

# 第 2 章　ベイズ理論

本章の構成

第 2 章第 1 節では、ベイズ理論の基となるベイズの定理を紹介し、この定理の証明をする。

第 2 章第 2 節では、ベイズの定理を用いることで得られる事後確率をもう一度、事前確率に置き直し、データから新たな事後確率を導くベイズ更新について説明し、一般化する。

第 2 章第 3 節では、事後確率と損失関数で表すことができるリスクを定義する。

第 2 章第 4 節では、意思決定にベイズ理論を利用した具体例を 2 つ紹介する。

また、第 2 章の内容は、既習事項として確率の基本的な性質、余事象、独立、条件付き確率、乗法定理、平均（期待値）を必要とするため、数学Ａの（場合の数と確率)、数学Ｂの（確率変数と確率分布）を終えていることが必要である。

## 第 1 節　ベイズの定理

今、$B$ を得られた結果、$A_1, A_2, \cdots, A_k$ を原因とする。私たちが知りたいのは、$B$ が起こったときの原因が $A_i$ である確率、すなわち $P(A_i|B)$ である。しかし、私たちが知ることができるのは原因に対する結果の確率 $P(B|A_i)$ である場合がほとんどであるから、$P(A_i|B)$ を直接計算することが困難な場合が多い。ベイズの定理は、結果に対する原因の確率 $P(A_i|B)$ を計算する公式を与える。

定理（ベイズの定理）

原因 $A_1, A_2, \cdots, A_k$ は互いに排反で、かつすべての場合をつくし、 $A_1 \cup A_2 \cup \cdots \cup A_k = \Omega$ とする。ここで、 $\Omega$ は全事象を表す。

このとき、

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{k} P(A_j)P(B|A_j)}$$

で計算される。ここに $P(A_i)$ は原因 $A_i$ の事前確率と呼ばれ、$P(A_i|B)$ は事後確率と呼ばれる。

（証明）

条件付確率の定義より

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

となる。
分子については、

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

より

$$P(A_i \cap B) = P(A_i)P(B|A_i)$$

となる。
分母については、$\{A_1 \cap B, A_2 \cap B, \cdots, A_k \cap B\}$ が互いに排反、かつ和が B であることに注意し、かつ上式の表現を再び用いれば、

$$P(B) = \sum_{j=1}^{k} P(A_j \cap B) = \sum_{j=1}^{k} P(A_j)P(B|A_j)$$

となる。すなわち、

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{k} P(A_j)P(B|A_j)}$$

を得る。

## 第2節　ベイズ更新

前節では、単一のデータに関する処理を考えてきた。次は、独立して得られる複数のデータに関する処理を考えていく。

前節から、結果の事象 $B$ が何度も生じる事象であると考えるときに、$t$ 回目の事象を $B_t(t = 1, 2, 3, \cdots)$ と表記する。前節では原因 $A_i(i = 1, 2, \cdots, k)$ というように $k$ 個の原因を考えていたが、ここでは、$A, A^c$ の 2 つの原因を考える。この時、原因 $A$ についてベイズの定理を使うと、

**1 回目**

$$\text{事前確率} = P(A)$$
$$\text{事後確率} = P(A|B_1) = \frac{P(A)P(B_1|A)}{P(A)P(B_1|A) + P(A^c)P(B_1|A^c)}$$

となる。

ここで、2 回目の事象 $B_2$ が生じたとする。すでに、事象 $B_1$ が生じているので、事前確率 $P(A)$ の代わりに、事後確率　$P(A|B_1)$ を新しい事前確率として使う。すると、事後確率は

**2 回目**

$$\text{事前確率} = P(A|B_1)$$
$$\text{事後確率} = P(A|B_1 \cap B_2) = \frac{P(A|B_1)P(B_2|A)}{P(A|B_1)P(B_2|A) + P(A^c|B_1)P(B_2|A^c)}$$

となる。

同様に繰り返し、$n$ 回目まで行うと、

**n 回目**

$$
\begin{aligned}
\text{事前確率} &= P(A|B_1 \cap B_2 \cap ... \cap B_{n-1}) \\
\text{事後確率} &= P(A|B_1 \cap B_2 \cap ... \cap B_n) \\
&= \frac{P(A|B_1 \cap B_2 \cap ... \cap B_{n-1})P(B_n|A)}{P(A|B_1 \cap B_2 \cap ... \cap B_{n-1})P(B_n|A) + P(A^c|B_1 \cap B_2 \cap ... \cap B_{n-1})P(B_n|A^c)}
\end{aligned}
$$

となる。

このように、事前確率に対して結果（データ）が得られ、ベイズの定理を用いることで事後確率を求めることができ、さらに、その事後確率を新たな事前確率として利用し、ベイズの定理を用いて、次の事後確率を求めるという処理を繰り返すことをベイズ更新と呼ぶ。

## 第 3 節　リスク

原因が $A_i(i = 1, \cdots, n)$ のときに行動 $a_k(k = 1, \cdots, l)$ とることによってこうむる損失関数を $L(A_i, a_k)$ とする。このとき、結果 $B_j(j = 1, \cdots, m)$[ここで、$j$ は事象の回数ではなく、事象の数、種類である] が得られた後での行動に対するリスク $R(B_j, a_k)$ は次のように定義される。

$$R(B_j, a_k) = \sum_{i=1}^{n} P(A_i|B_j)L(A_i, a_k)$$

また、結果の事象が何度も得られる事象であるとき、ベイズ更新によって、何度も事後確率は更新されていくが、上式の事後確率には、その更新された事後確率を利用する。

そして、各結果におけるリスクを比較したときに、最小のリスクを導く行動をとることが最良の意思決定となる。つまり、

$$\min\{R(B_j, a_1), R(B_j, a_2), \cdots, R(B_j, a_l)\}$$

となる行動 $a$ をとる。

## 第 4 節　具体例

例 1

あるお弁当屋は明日、運動会があるため、天気予報を聞いてからお弁当を用意することにした。この予報の信頼性を調べたところ、過去のデータから次の確率が得られた。

| | | 明日の天気 | |
|---|---|---|---|
| | | 晴れまたは曇り | 雨 |
| 天気予報 | 「晴れまたは曇り」と発表 | $\frac{7}{10}$ | $\frac{2}{10}$ |
| | 「雨」と発表 | $\frac{3}{10}$ | $\frac{8}{10}$ |

表 1: 天気予報の信頼性の確率

例えば左上の $\frac{7}{10}$ とは、明日の天気が実際に晴れまたは曇りのときに、天気予報が正しく「晴れまたは曇り」と発表する確率である。誤って「雨」と発表する確率はその下の $\frac{3}{10}$ である。

さて、お弁当屋の行動として次の 3 つを考える。

$a_1$ : お弁当を 100 個用意する

$a_2$ : お弁当を 70 個用意する

$a_3$ : お弁当を 50 個用意する

このとき、条件として、晴れまたは曇りの場合はお弁当が全て売れ、雨の場合はお弁当が用意した分の 2 割しか売れない。また、お弁当の金額は 1 個 500 円で、原価 300 円としたとき、損失関数を次のように定義する。

$$損失関数 = \begin{cases} 20000 - (200 \times 用意したお弁当の数) & (晴れまたは曇りの場合) \\ 300 \times 残ったお弁当の数 - 200 \times 売れたお弁当の数 & (雨の場合) \end{cases}$$

ここで、一般に、損失関数を 0 または正の値をとる実数値関数と定義し、式で表すことがほとんどないため、生徒の理解が進まないと考え、上式のように定義する。上式の晴れまたは曇りの場合における 20000 は、お弁当は最大で 100 個売れるため、そのときの利益である 20000 円を表している。そして、その 20000 から実際に用意したお弁当の数の時の利益を引いたものが式となっている。例えば、晴れまたは曇りの時に、行動 $a_2$（お弁当を 70 個用意する）をとると、損失は $20000 - 200 \times 70 = 20000 - 14000 = 6000$ となる。

また、雨の場合には、それぞれの行動をとったときの損失額を表している。例えば、雨の時に、行動 $a_2$（お弁当を 70 個用意する）をとると、損失は $300 \times 56 - 200 \times 14 = 16800 - 2800 = 14000$ となる。

すると、損失表は以下のようになる。

| | | 明日の天気 | |
|---|---|---|---|
| | | 晴れまたは曇り | 雨 |
| 行動 | $a_1$ | 0 | 20000 |
| | $a_2$ | 6000 | 14000 |
| | $a_3$ | 10000 | 10000 |

表 2: 行動に対する損失表

一番左上は、明日の天気が「晴れまたは曇り」のときに行動 $a_1$ をとった時の損失である。

では、天気予報で「明日の天気は晴れまたは曇りです。」と言われたお弁当屋は、どの行動をするのがよいのか。また、「明日の天気は雨です。」と言われたお弁当屋は、どの行動をするのがよいのか。

ここで、事前データとして、過去のデータから 10 日間の天気は、次のようであったとする。

| | 晴れまたは曇り | 雨 |
|---|---|---|
| 10 日間 | 7 日 | 3 日 |

表 3: 10 日間の天気

まず、事象をアルファベットを用いて表記する。
原因
$A$ : 実際の天気が「晴れまたは曇り」
$A^c$ : 実際の天気が「雨」
結果 (データ)
$B$ : 天気予報が「晴れまたは曇り」と発表
$B^c$ : 天気予報が「雨」と発表

次に、ベイズの定理の特徴である事前確率を設定する。表 3 に過去のデータから 10 日間の天気が掲載されているので、これを事前確率して利用する。
事前確率は

$$P(\text{実際の天気が「晴れまたは曇り」}) = P(A) = \frac{7}{10}$$
$$P(\text{実際の天気が「雨」}) = P(A^c) = \frac{3}{10}$$

となる。
また、表1から予報の信頼性の確率は

$P($天気予報が「晴れまたは雲り」と発表 | 実際の天気が「晴れまたは曇り」$)$
$= P(B|A) = \dfrac{7}{10}$
$P($天気予報が「晴れまたは雲り」と発表 | 実際の天気が「雨」$)$
$= P(B|A^c) = \dfrac{3}{10}$
$P($天気予報が「雨」と発表 | 実際の天気が「晴れまたは曇り」$)$
$= P(B^c|A) = \dfrac{2}{10}$
$P($天気予報が「雨」と発表 | 実際の天気が「雨」$)$
$= P(B^c|A^c) = \dfrac{8}{10}$

である。
　ここで、ベイズの定理を用いて、事後確率を算出すると、

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)} \\ &= \frac{(\frac{7}{10})(\frac{7}{10})}{(\frac{7}{10})(\frac{7}{10})+(\frac{3}{10})(\frac{2}{10})} = 0.891 \end{aligned}$$

$$\begin{aligned} P(A|B^c) &= \frac{P(A)P(B^c|A)}{P(A)P(B^c|A)+P(A^c)P(B^c|A^c)} \\ &= \frac{(\frac{7}{10})(\frac{3}{10})}{(\frac{7}{10})(\frac{3}{10})+(\frac{3}{10})(\frac{8}{10})} = 0.467 \end{aligned}$$

$$\begin{aligned} P(A^c|B) &= \frac{P(A^c)P(B|A^c)}{P(A)P(B|A)+P(A^c)P(B|A^c)} \\ &= \frac{(\frac{3}{10})(\frac{2}{10})}{(\frac{7}{10})(\frac{7}{10})+(\frac{3}{10})(\frac{2}{10})} = 0.109 \end{aligned}$$

$$\begin{aligned} P(A^c|B^c) &= \frac{P(A^c)P(B^c|A^c)}{P(A)P(B^c|A)+P(A^c)P(B^c|A^c)} \\ &= \frac{(\frac{3}{10})(\frac{8}{10})}{(\frac{7}{10})(\frac{3}{10})+(\frac{3}{10})(\frac{8}{10})} = 0.533 \end{aligned}$$

である。

従って、事後確率は

$P$(実際の天気が「晴れまたは曇り」| 天気予報が「晴れまたは曇り」と発表)

$= P(A|B) = 0.891$

$P$(実際の天気が「晴れまたは曇り」| 天気予報が「雨」と発表)

$= P(A|B^c) = 0.467$

$P$(実際の天気が「雨」| 天気予報が「晴れまたは曇り」と発表)

$= P(A^c|B) = 0.109$

$P$(実際の天気が「雨」| 天気予報が「雨」と発表)

$= P(A^c|B^c) = 0.533$

となる。

ここで、行動に対するリスクを求める。

前節の式を用いて、天気予報が「晴れまたは曇り」と発表したとき、行動 $a_1$ のリスクを算出すると

$$R(B, a_1) = 0.891 \times 0 + 0.109 \times 20000 = 2180$$

となる。あと 2 つの行動のリスクについても同様に求めると

$$R(B, a_2) = 0.891 \times 6000 + 0.109 \times 14000 = 6436$$
$$R(B, a_3) = 0.891 \times 10000 + 0.109 \times 10000 = 10000$$

となる。また、天気予報が「雨」と発表したときの各行動のリスクは

$$R(B^c, a_1) = 0.467 \times 0 + 0.533 \times 20000 = 10660$$
$$R(B^c, a_2) = 0.467 \times 6000 + 0.533 \times 14000 = 10264$$
$$R(B^c, a_3) = 0.467 \times 10000 + 0.533 \times 10000 = 10000$$

となる。これらのリスクを得られたことにより、次の表が得られる。

| | | 行動 | | |
|---|---|---|---|---|
| | | $a_1$ | $a_2$ | $a_3$ |
| 天気予報 | 「晴れまたは曇り」と発表 | 2180 | 6436 | 10000 |
| | 「雨」と発表 | 10660 | 10264 | 10000 |

表 4: 行動に対するリスク表

「『晴れまたは曇り』と発表」されたときにリスクを最小にする行動は行動 $a_1$(お弁当を 100 個用意する) であり、「『雨』と発表」されたときにリスクを最小にする行動は行動 $a_3$(お弁当を 50 個用意する) である。こうして、次のように解答が得られる。

| | | 最小のリスクを期待できる行動 |
|---|---|---|
| 天気予報 | 「晴れまたは曇り」と発表 | 行動 $a_1$(お弁当を 100 個用意する) |
| | 「雨」と発表 | 行動 $a_3$(お弁当を 50 個用意する) |

表 5: 明日ためにとるべき行動

次に、午前の天気予報と午後の天気予報の 2 回、天気予報を聞いてからお弁当を用意すると考える。
午前の天気予報を
$B_1$：午前の天気予報が「晴れまたは曇り」と発表する
$B_1^c$：午前の天気予報が「雨」と発表する
と表記する。
このときに、午後の天気予報は
$B_2$：午後の天気予報が「晴れまたは曇り」と発表する
$B_2^c$：午後の天気予報が「雨」と発表する
とする。そして、午前、午後の予報の信頼性の確率は表 1 で与えられる条件付き確率で互いに独立に生じるとする。
この場合、2 回天気予報を聞くことができるため、ベイズ更新を使うことができる。
従って、まず、午前の天気予報で「晴れまたは曇り」と発表した場合を考える。このとき、1 回目で求めた事後確率を事前確率として用いる。
つまり、事前確率を

$$P(A|B_1) = 0.891, \quad P(A^c|B_1) = 0.109$$

とする。
ここで、ベイズの定理を用いて、事後確率を算出すると、

$$\begin{aligned} P(A|B_1 \cap B_2) &= \frac{P(A|B_1)P(B_2|A)}{P(A|B_1)P(B_2|A) + P(A^c|B_1)P(B_2|A^c)} \\ &= \frac{0.891 \times (\frac{7}{10})}{0.891 \times (\frac{7}{10}) + 0.109 \times (\frac{2}{10})} = 0.966 \end{aligned}$$

$$\begin{aligned} P(A|B_1 \cap B_2^c) &= \frac{P(A|B_1)P(B_2^c|A)}{P(A|B_1)P(B_2^c|A) + P(A^c|B_1)P(B_2^c|A^c)} \\ &= \frac{0.891 \times (\frac{3}{10})}{0.891 \times (\frac{3}{10}) + 0.109 \times (\frac{8}{10})} = 0.754 \end{aligned}$$

$$
\begin{aligned}
P(A^c|B_1\cap B_2) &= \frac{P(A^c|B_1)P(B_2|A^c)}{P(A|B_1)P(B_2|A)+P(A^c|B_1)P(B_2|A^c)} \\
&= \frac{0.109\times(\frac{2}{10})}{0.891\times(\frac{7}{10})+0.109\times(\frac{2}{10})} = 0.034
\end{aligned}
$$

$$
\begin{aligned}
P(A^c|B_1\cap B_2^c) &= \frac{P(A^c|B_1)P(B_2^c|A^c)}{P(A|B_1)P(B_2^c|A)+P(A^c|B_1)P(B_2^c|A^c)} \\
&= \frac{0.109\times(\frac{8}{10})}{0.891\times(\frac{3}{10})+0.109\times(\frac{8}{10})} = 0.246
\end{aligned}
$$

である。
従って、事後確率は、

$P$(実際の天気が「晴れまたは曇り」| 午前「晴れまたは曇り」$\cap$ 午後「晴れまたは曇り」)

$= P(A|B_1\cap B_2) = 0.966$

$P$(実際の天気が「晴れまたは曇り」| 午前「晴れまたは曇り」$\cap$ 午後「雨」)

$= P(A|B_1\cap B_2^c) = 0.754$

$P$(実際の天気が「雨」| 午前「晴れまたは曇り」$\cap$ 午後「晴れまたは曇り」)

$= P(A^c|B_1\cap B_2) = 0.034$

$P$(実際の天気が「雨」| 午前「晴れまたは曇り」$\cap$ 午後「雨」)

$= P(A^c|B_1\cap B_2^c) = 0.246$

となる。
　また、同様に、午前の天気予報が「雨」と発表した場合を考えると、資料 1 のように事後確率は、

$P$(実際の天気が「晴れまたは曇り」| 午前「雨」$\cap$ 午後「晴れまたは曇り」)

$= P(A|B_1^c\cap B_2) = 0.754$

$P$(実際の天気が「晴れまたは曇り」| 午前「雨」$\cap$ 午後「雨」)

$= P(A|B_1^c\cap B_2^c) = 0.247$

$P$(実際の天気が「雨」| 午前「雨」$\cap$ 午後「晴れまたは曇り」)

$= P(A^c|B_1^c\cap B_2) = 0.246$

$P$(実際の天気が「雨」| 午前「雨」$\cap$ 午後「雨」)

$= P(A^c|B_1^c\cap B_2^c) = 0.753$

となる。これらの事後確率より、以下の表が得られる。

| | | 午前と午後の天気予報 | | | |
|---|---|---|---|---|---|
| | | $B_1 \cap B_2$ | $B_1 \cap B_2^c$ | $B_1^c \cap B_2$ | $B_1^c \cap B_2^c$ |
| 実際の天気 | $A$ | 0.966 | 0.754 | 0.754 | 0.247 |
| | $A^c$ | 0.034 | 0.246 | 0.246 | 0.753 |

表 6: 2 回天気予報を聞いたときの事後確率

1 回目の事前確率から始めて、天気予報が「晴れまたは曇り」と発表する回数が 1 回、2 回と増えるにつれ、実際の天気が「晴れまたは曇り」である事後確率がどのように変化するかをまとめると、表 7 のようになる。

| | 明日の天気 | |
|---|---|---|
| | 晴れまたは曇り ($A$) | 雨 ($A^c$) |
| 事前確率 | 0.7 | 0.3 |
| 事後確率 (1 回目) | 0.891 | 0.109 |
| 事後確率 (2 回目) | 0.966 | 0.034 |

表 7: 「晴れまたは曇り」と発表したときの事後確率

表 7 では、最初の事前確率から、天気予報が「晴れまたは曇り」と発表するたびに、明日の天気が「晴れまたは曇り」である事後確率が高くなっていくことがわかる。

ここで、もう 1 度、表 6 を見てみる。午前「晴れまたは曇り」、午後「雨」と発表した場合と午前「雨」、午後「晴れまたは曇り」と発表した場合の 2 つの事後確率を比べると、同じ確率になっている。ベイズ更新では、原因に対する 1 回目の結果と 2 回目の結果を順番に処理した場合の事後確率と同時に 2 つ処理した場合の事後確率が等しくなる。そのため、結果を処理する順序に関係なく、各結果が得られる回数が同じであれば、同じ事後確率になる。このような性質を、逐次合理性という。例 1 の場合、「晴れまたは曇り」を 1 回、「雨」を 1 回、聞いた場合の事後確率が等しくなっている。さらに、同時に処理した場合の事後確率は、条件付き独立性 $P(B \cap B^c|A) = P(B|A)P(B^c|A)$ より

$$\begin{aligned} P(A|B \cap B^c) &= \frac{P(A)P(B \cap B^c|A)}{P(A)P(B \cap B^c|A) + P(A^c)P(B \cap B^c|A^c)} \\ &= \frac{P(A)P(B|A)P(B^c|A)}{P(A)P(B|A)P(B^c|A) + P(A^c)P(B|A^c)P(B^c|A^c)} \\ &= \frac{(\frac{7}{10})(\frac{7}{10})(\frac{3}{10})}{(\frac{7}{10})(\frac{7}{10})(\frac{3}{10}) + (\frac{3}{10})(\frac{2}{10})(\frac{8}{10})} = 0.754 \end{aligned}$$

$$
\begin{aligned}
P(A^c|B\cap B^c) &= \frac{P(A^c)P(B\cap B^c|A^c)}{P(A)P(B\cap B^c|A)+P(A^c)P(B\cap B^c|A^c)} \\
&= \frac{P(A^c)P(B|A^c)P(B^c|A^c)}{P(A)P(B|A)P(B^c|A)+P(A^c)P(B|A^c)P(B^c|A^c)} \\
&= \frac{(\frac{3}{10})(\frac{2}{10})(\frac{8}{10})}{(\frac{7}{10})(\frac{7}{10})(\frac{3}{10})+(\frac{3}{10})(\frac{2}{10})(\frac{8}{10})} = 0.246
\end{aligned}
$$

となり、2 回に分けて処理した事後確率と等しいことがわかる。

次に、1 回しか天気予報を聞かない場合と同じように、行動に対するリスクを求める。

午前の天気予報が「晴れまたは曇り」、午後の天気予報も「晴れまたは曇り」と発表したときのリスクを算出すると、

$$
\begin{aligned}
R(B_1\cap B_2, a_1) &= 0.966\times 0 + 0.034\times 20000 = 680 \\
R(B_1\cap B_2, a_2) &= 0.966\times 6000 + 0.034\times 14000 = 6272 \\
R(B_1\cap B_2, a_3) &= 0.966\times 10000 + 0.034\times 10000 = 10000
\end{aligned}
$$

となる。また、他の場合も資料 1 のように求められ、次の表が得られる。

| | | 行動 | | |
|---|---|---|---|---|
| | | $a_1$ | $a_2$ | $a_3$ |
| 午前と午後の天気予報 | $B_1\cap B_2$ | 680 | 6272 | 10000 |
| | $B_1\cap B_2^c$ | 4920 | 7968 | 10000 |
| | $B_1^c\cap B_2$ | 4920 | 7968 | 10000 |
| | $B_1^c\cap B_2^c$ | 15060 | 12024 | 10000 |

表 8: 行動に対するリスク表

よって、リスクを最小にする行動は、以下の通りである。

- 午前の天気予報が「晴れまたは曇り」、午後の天気予報が「晴れまたは曇り」と発表したとき、行動 $a_1$(お弁当を 100 個用意する) をとる
- 午前の天気予報が「晴れまたは曇り」、午後の天気予報が「雨」と発表したとき、行動 $a_1$(お弁当を 100 個用意する) をとる
- 午前の天気予報が「雨」、午後の天気予報が「晴れまたは曇り」と発表したとき、行動 $a_1$(お弁当を 100 個用意する) をとる
- 午前の天気予報が「雨」、午後の天気予報が「雨」と発表したとき、行動 $a_3$(お弁当を 50 個用意する) をとる

**例 2**(中妻,2007)

企業に対する信頼がどのように得られるか（あるいは失われるか）を考える。融資に限らず企業と取引をするうえで相手企業に信頼できるかどうかは非常に重要である。そして、信頼できる企業である確率が取引先企業の行動によって更新される仕組みを満たすことで、ベイズ理論の基礎であるベイズの定理が理解できる。

では、どのようにして相手が信頼できる企業かどうかを見分けるのか。信頼できる企業であるかどうかを過去の企業の行動から判断することにする。ここでは、納期をきちんと守ったかどうかを判断材料として企業の信頼度を推測することを考えていく。

ここで、取引先の企業をX社とする。これからX社と取引をしようとする企業の担当者にとって、X社が信頼できる企業か信頼できない企業かを事前に判別できないとする。X社の信頼度は前もって観測できないものの、X社が信頼できる企業であればきちんと納期を守り、信頼できない場合には納期に遅れがちになるとする。そうすると、X社が過去にどれだけ納期を守ったかをみることで観測できないX社の信頼度を推測できそうである。例えば、納期を必ず守っていれば信頼できると判断し、一度でも納期に遅れれば信頼できないと判断する、という基準でX社の信頼度を推測することができる。しかし、これは少々厳し過ぎる基準かもしれない。なぜなら、納期の遅れはX社に責任のない何らかの不可抗力によるものかもしれないからである。従って、一度納期に遅れたからといって、X社を全く信頼できない企業であると決めつけるのは早計である。しかし、あまりにも頻繁に納期を守らないとX社は信頼できないと判断せざるを得ない。

では、何回納期に遅れたらX社を信頼できない企業であると判断すればよいのか。まず、事象をアルファベットを用いて表記する。

原因

$A$ : X社は信頼できる

$A^c$ : X社は信頼できない

結果

$B$ : 納期に遅れる

$B^c$ : 納期を守る（これは、納期に遅れないことである）

次にもし、X社が信頼できる企業であれば納期に遅れることはないとする。これは「X社が信頼できる企業である」という事象 $B$ が起きれば「納期に遅れる」という事象は $A$ は起きないということを意味する。つまり、X社が信頼できる企業であるという条件の下で納期に遅れる条件付き確率 $P(B|A)$ は

$$P(B|A) = 0$$

である。逆にX社が信頼できる企業であるという条件の下で納期を守る条件付き確

率 $P(B^c|A)$ は

$$P(B^c|A) = 1$$

となる。条件付き確率も確率であることに変わりはないので $P(B|A) + P(B^c|A) = 1$ が成り立つ。

一方、X社が信頼できない企業である場合には2回に1回の割合で納期に遅れると仮定する。このときX社が信頼できない企業であるという条件の下で納期に遅れる条件付き確率 $P(B|A^c)$ と納期を守る条件付き確率 $P(B^c|A^c)$ は、それぞれ

$$P(B|A^c) = \frac{1}{2}, \quad P(B^c|A^c) = \frac{1}{2}$$

である。ここでも $P(B|A^c)$ と $P(B^c|A^c)$ の和は1である。以上をまとめると納期の遅れに関する条件付き確率は表9のようになる。

| | 信頼できる $(A)$ | 信頼できない $(A^c)$ |
|---|---|---|
| 納期に遅れる $(B)$ | 0 | $\frac{1}{2}$ |
| 納期を守る $(B^c)$ | 1 | $\frac{1}{2}$ |

表 9: 納期の遅れの条件付き確率（その１）

それではX社の信頼度の推測にベイズの定理がどのように使われるかをみてみる。ここでX社の取引相手がX社を信頼できる企業かどうか五分五分であると考えていると仮定する。すると、事前確率は

$$P(A) = P(A^c) = \frac{1}{2} \tag{1}$$

である。

さて、実際にX社が納期を守らなかったとする。ベイズの定理を使うと、X社が納期を守らなかったという条件の下でX社が信頼できる企業である条件付き確率は

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} \\ &= \frac{(\frac{1}{2}) \times 0}{(\frac{1}{2}) \times 0 + (\frac{1}{2}) \times (\frac{1}{2})} = \frac{0}{\frac{1}{4}} = 0 \end{aligned} \tag{2}$$

となる。これが事後確率である。事後確率がゼロであるから、納期を守らなかったX社は信頼できる企業である可能性はない、つまり信頼できないと結論づけることになる。この例では「納期に遅れた」という新しい情報によってX社が信頼できる企業である可能性が五分五分からゼロへ一気に下がってしまったが、一度納期に遅れたからといって信頼できないと決めつけるのも厳しいようにみえるかもしれない。しかし、表9では信頼できる企業は必ず納期を守ると想定しているので、一度でも

納期に遅れれば即信頼できない企業と断定することができる。

では何らかの手違いで信頼できる企業でも納期に遅れることがあるとしてみる。例えば、信頼できる企業の納期に関する条件付き確率が

$$P(B|A) = \frac{1}{20}, \quad P(B^c|A) = \frac{19}{20} \tag{3}$$

であると仮定する。これは信頼できる企業でも 20 回に 1 回の割合で納期に遅れてしまうという状況を想定している。他の確率はすべて同じであるとすると、納期の遅れの条件付き確率の表 10 のようになる。

| | 信頼できる $(A)$ | 信頼できない $(A^c)$ |
|---|---|---|
| 納期に遅れる $(B)$ | $\frac{1}{20}$ | $\frac{1}{2}$ |
| 納期を守る $(B^c)$ | $\frac{19}{20}$ | $\frac{1}{2}$ |

表 10: 納期の遅れの条件付き確率（その 2）

この場合X社が納期を守らなかったときにX社が信頼できる企業である事後確率は

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)} \\ &= \frac{(\frac{1}{2})(\frac{1}{20})}{(\frac{1}{2})(\frac{1}{20})+(\frac{1}{2})(\frac{1}{2})} = \frac{\frac{1}{40}}{\frac{11}{40}} = \frac{1}{11} \end{aligned} \tag{4}$$

となる。この例では、一度納期に遅れただけで信頼できる企業である事後確率がいきなりゼロになるということはないが、それでも $\frac{1}{2}$ から $\frac{1}{11}$ へと急落していく。

次に、信頼できる企業の納期に関する条件付き確率は引き続いて (3) 式のものだが、事前確率が

$$P(A) = 1, \quad P(A^c) = 0 \tag{5}$$

である場合を考える。これはX社が信頼できる企業であると絶対的に確信している状況を表している。すると、X社が納期を守らなかったときにX社が信頼できる企業である事後確率は

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)} \\ &= \frac{1\times(\frac{1}{20})}{1\times(\frac{1}{20})+0\times(\frac{1}{2})} = \frac{\frac{1}{20}}{\frac{1}{20}} = 1 \end{aligned} \tag{6}$$

となる。つまり、事前確率が (5) 式で与えられる場合にはX社が信頼できる企業であると確信しているため、「Xが納期に遅れた」という新しい情報が入ってきてもX社に対する信頼度が全く揺るがないのである。

今までは1回だけ納期の遅れた場合のX社が信頼できる企業である事後確率をみてきた。次に、X社が続けて何回も納期に遅れた場合にX社に対する信頼がどのように失われいていくかをみていく。「$i$回目の納期に遅れる」という事象を$B_i (i = 1, 2, 3, ...)$と表記する。そして、各回の納期の遅れは表10で与えられる同じ条件付き確率で互いに独立に生じるとする。ここで2回目の納期の遅れが生じたとする。すでにX社が一度納期に遅れたことはわかっているので、(1)式の事前分布の代わりに(4)式の事後確率を新しい事前確率

$$P(A|B_1) = \frac{1}{11}, \quad P(A^c|B_1) = \frac{10}{11}$$

に使う。すると、X社が信頼できる企業である事後確率は

$$\begin{aligned} P(A|B_1 \cap B_2) &= \frac{P(A|B_1)P(B_2|A)}{P(A|B_1)P(B_2|A) + P(A^c|B_1)P(B_2|A^c)} \\ &= \frac{(\frac{1}{11})(\frac{1}{20})}{(\frac{1}{11})(\frac{1}{20}) + (\frac{10}{11})(\frac{1}{2})} = \frac{\frac{1}{220}}{\frac{101}{220}} = \frac{1}{101} \end{aligned} \tag{7}$$

となる。(7)式の$P(A|B_1 \cap B_2)$は、「1回目と2回目の納期に遅れる」という事象$B_1 \cap B_2$が起きたという条件の下で「X社が信頼できる企業とわかる」という事象$A$が起きる条件付き確率である。さらに3回目の納期の遅れが生じたとする。今度は(7)式の事後確率を新しい事前確率

$$P(B|A_1 \cap A_2) = \frac{1}{101}, \quad P(B^c|A_1 \cap A_2) = \frac{100}{101}$$

に使うと、X社が信頼できる企業である事後確率は

$$\begin{aligned} P(A|B_1 \cap B_2 \cap B_3) &= \frac{P(A|B_1 \cap B_2)P(B_3|A)}{P(A|B_1 \cap B_2)P(B_3|A) + P(A^c|B_1 \cap B_2)P(B_3|A^c)} \\ &= \frac{(\frac{1}{101})(\frac{1}{20})}{(\frac{1}{101})(\frac{1}{20}) + (\frac{100}{101})(\frac{1}{2})} = \frac{\frac{1}{2020}}{\frac{1001}{2020}} = \frac{1}{1001} \end{aligned} \tag{8}$$

となる。(1)式の事前確率から始めて、納期の遅れが1回、2回、3回と増えるにつれX社が信頼できる企業である事後確率がどのように変化するかをまとめると、表11のようになる。

| | 信頼できる ($A$) | 信頼できない ($A^c$) |
|---|---|---|
| 事前確率 | $\frac{1}{2}$ | $\frac{1}{2}$ |
| 事後確率 (1回目) | $\frac{1}{11}$ | $\frac{10}{11}$ |
| 事後確率 (2回目) | $\frac{1}{101}$ | $\frac{100}{101}$ |
| 事後確率 (3回目) | $\frac{1}{1001}$ | $\frac{1000}{1001}$ |

表11: 納期の遅れの条件付き確率（その2）

表 11 では、最初の事前確率の段階ではX社は五分五分で信頼できると考えていたが、納期に遅れるたびにX社が信頼できる企業である事後確率は低下し続ける。そして、3 回も納期に遅れた後ではX社が信頼できる企業である事後確率は 0.1%を下回ってしまっている。それだけX社に対する信頼が失われたということである。

ちなみに (5) 式の事前確率を使って同じことをしてみると

$$
\begin{aligned}
P(A|B_1) &= \frac{P(A)P(B_1|A)}{P(A)P(B_1|A)+P(A^c)P(B_1|A^c)} \\
&= \frac{1\times(\frac{1}{20})}{1\times(\frac{1}{20})+0\times(\frac{1}{2})} = \frac{\frac{1}{20}}{\frac{1}{20}} = 1
\end{aligned}
$$

$$
\begin{aligned}
P(A|B_1\cap B_2) &= \frac{P(A|B_1)P(B_2|A)}{P(A|B_1)P(B_2|A)+P(A^c|B_1)P(B_2|A^c)} \\
&= \frac{1\times(\frac{1}{20})}{1\times(\frac{1}{20})+0\times(\frac{1}{2})} = \frac{\frac{1}{20}}{\frac{1}{20}} = 1
\end{aligned}
$$

$$
\begin{aligned}
P(A|B_1\cap B_2\cap B_3) &= \frac{P(A|B_1\cap B_2)P(B_3|A)}{P(A|B_1\cap B_2)P(B_3|A)+P(A^c|B_1\cap B_2)P(B_3|A^c)} \\
&= \frac{1\times(\frac{1}{20})}{1\times(\frac{1}{20})+0\times(\frac{1}{2})} = \frac{\frac{1}{20}}{\frac{1}{20}} = 1
\end{aligned}
$$

となる。さらにX社 4 回、5 回、6 回と何度納期に遅れてもX社が信頼できる企業である事後確率はずっと 1 のままである。つまり、X社が信頼できる企業であると盲信しているため、何度裏切られても信頼度は揺るがないのである。

# 第 3 章　統計的決定問題

本章の構成

第 3 章第 1 節では、決定理論を紹介する。

第 3 章第 2 節では、決定方式を選ぶ際の順序付けと許容性について定義する。

第 3 章第 3 節では、意思決定方式であるミニマックス決定方式とベイズ決定方式の 2 つを定義し、ベイズ決定方式から得られる事後期待損失最小化の定理を証明する。

## 第 1 節　決定理論とは

次の (1)〜(6) を決定理論と呼ぶ。　(1) 自然の状態 $\theta$ :考えうる $\theta$ の全体を $\Theta$ とおき、状態の空間と呼ぶ。（第 2 章の例では、$A$ にあたる）

(2) 行動 $a$: 考えうる $a$ 全体を $A$ とおき、行動空間と呼ぶ。（第 2 章の例と同じ）

(3) 損失関数 $L(\theta, a)$ : 正または 0 の値をとる。ただし、負としても矛盾はおきない。（第 2 章では、$L(A, a)$ にあたる）

$a$ を$\theta$ の推定値とするときにも損失関数を使うが、その時の代表的なものとしては、

$$\text{絶対損失}\quad L(\theta, a) = |a - \theta|$$

$$\text{二次損失}\quad L(\theta, a) = \lambda(\theta)(a - \theta)^2$$

$$0 - 1\,\text{型損失}\quad L(\theta, a) = \begin{cases} 0 & (|a - \theta| \leq \Delta) \\ 1 & (|a - \theta| > \Delta) \end{cases}$$

の 3 つである。

(4) データ $z$:必ずしも数量化されていない。また、数量であっても 1 次元とは限らない。$z$ 全体（ $L$ 通り）を Ł とおき、標本空間と呼ぶ。（第 2 章では、$B$ にあたる）

(5) データの確率分布 $p(z|\theta)$。（第 2 章の例では、$P(B|A)$ にあたる）

(6) 決定方式 $d : z \to a$。　数学的には Ł から $A$ への写像、その全体を $D$ で表す。

具体的な例では、次のようになる。

例

第 2 章の具体例の例 1 を決定理論に用いる。

このとき、

$$\text{状態}\begin{cases} \theta_1 : \text{明日の天気が「晴れまたは曇り」。} \\ \theta_2 : \text{明日の天気が「雨」。} \end{cases}$$

$$
\text{行動}\begin{cases} a_1：\text{お弁当を 100 個用意する。} \\ a_2：\text{お弁当を 70 個用意する。} \\ a_3：\text{お弁当を 50 個用意する。} \end{cases}
$$

$$
\text{データ}\begin{cases} z_1：\text{天気予報が「晴れまたは曇り」と発表。} \\ z_2：\text{天気予報が「雨」と発表。} \end{cases}
$$

また、損失関数も次のように定義する。

$$
\text{損失関数} = \begin{cases} 20000 - (200 \times \text{用意したお弁当の数}) & (\text{状態が}\theta_1\text{の場合}) \\ 300 \times \text{残ったお弁当の数} - 200 \times \text{売れた数} & (\text{状態が}\theta_2\text{場合}) \end{cases}
$$

すると、損失表とデータの確率分布は以下のようになる。

| | | 状態 | |
|---|---|---|---|
| | | $\theta_1$ | $\theta_2$ |
| 行動 | $a_1$ | 0 | 20000 |
| | $a_2$ | 6000 | 14000 |
| | $a_3$ | 10000 | 10000 |

表 12: 行動に対する損失表

| | | 状態 | |
|---|---|---|---|
| | | $\theta_1$ | $\theta_2$ |
| データ | $z_1$ | $\frac{7}{10}$ | $\frac{2}{10}$ |
| | $z_2$ | $\frac{3}{10}$ | $\frac{8}{10}$ |

表 13: データの確率分布

| | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_7$ | $d_8$ | $d_9$ |
|---|---|---|---|---|---|---|---|---|---|
| $z_1$ | $a_1$ | $a_1$ | $a_1$ | $a_2$ | $a_2$ | $a_2$ | $a_3$ | $a_3$ | $a_3$ |
| $z_2$ | $a_1$ | $a_2$ | $a_3$ | $a_1$ | $a_2$ | $a_3$ | $a_1$ | $a_2$ | $a_3$ |

表 14: 全ての決定方式のリスト

また、データ $z_i$ が観測されたとき、行動 $a_j$ をとるということを $d(z_i) = a_j$ とすると、表 14 における $d_2$ は、$z_1$ が観測されたときには行動 $a_1$ を、$z_2$ が得られたときには行動 $a_2$ をとるような決定方式、つまり、$d_2(z_1) = a_1$、$d_2(z_2) = a_2$ を表ししている。

## 第 2 節　意思決定方式を選択するための順序付けと許容性

表 14 のそれぞれの決定方式は適切な行動 $a$ を指示することもあるが、そうでない行動を指示することもありうる。この程度のことを一般リスクと呼ぶ。決定方式 $d$ の一般リスクは、表 14 に見るように、その度合は真の状態 $\theta$ が $\theta_1$ か $\theta_2$ かによっ

て異なる。従って $\theta$ の関数でもあり、$R(\theta, d)$ と表す。そこで、$d$ の一般リスクを、損失の期待値

$$R(\theta_i, d) = \sum_{l=1}^{L} p(z_l|\theta_i) L(\theta_i, d(z_l))$$

と定義しよう。とりうるデータの値を表す確率変数を $\tilde{z}$ とするとき、$d$ によって引き起こされる損失 $L(\theta, d(\tilde{z}))$ の期待値が $d$ の一般リスクなのである。各点をリスク点といい、またリスク点の集合をリスク集合といい、$R$ で表す。先ほどの例の場合、決定方式 $d_1$ の状態 $\theta_1$ に対する一般リスク $R(\theta_1, d_1)$ を求めると、

$$\begin{aligned} R(\theta_1, d_1) &= p(z_1|\theta_1)L(\theta_1, d_1(z_1)) + p(z_2|\theta_1)L(\theta_1, d_1(z_2)) \\ &= p(z_1|\theta_1)L(\theta_1, a_1) + p(z_2|\theta_1)L(\theta_1, a_1) \\ &= \frac{7}{10} \times 0 + \frac{2}{10} \times 0 = 0 \end{aligned}$$

である。また、$R(\theta_2, d_1)$ を求めると、

$$\begin{aligned} R(\theta_2, d_1) &= p(z_1|\theta_2)L(\theta_2, d_1(z_1)) + p(z_2|\theta_2)L(\theta_2, d_1(z_2)) \\ &= p(z_1|\theta_2)L(\theta_2, a_1) + p(z_2|\theta_2)L(\theta_2, a_1) \\ &= \frac{3}{10} \times 20000 + \frac{8}{10} \times 20000 = 22000 \end{aligned}$$

である。これを $d_2$〜$d_9$ まで同様に続けると、資料 2 のようになる。

　これらより、各決定方式の一般リスクは表 15 にまとめられる。

| | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_7$ | $d_8$ | $d_9$ |
|---|---|---|---|---|---|---|---|---|---|
| $R(\theta_1, d)$ | 0 | 1200 | 2000 | 4200 | 5400 | 6200 | 7000 | 8200 | 9000 |
| $R(\theta_2, d)$ | 22000 | 17200 | 14000 | 20200 | 15400 | 12200 | 19000 | 14200 | 11000 |

表 15: 各決定方式の一般リスク

一般に、2 つの決定方式 $d, d'$ をとった場合、その一般リスクには次のような関係が考えられる（状態の数=2 のとき）。一般に、比較して

$$\text{すべての}\theta\text{に対し、}R(\theta, d) \leq R(\theta, d') \qquad (*)$$

が成り立てば、$d$ は $d'$ よりベターか、あるいは同じ良さの程度（つまり、少なくとも同じ良さの程度）をもつことになる。

　この比較が本質的となるためには、式 $(*)$ を少し修正することが望ましい。なぜなら、$d$ と $d'$ が偶然にも全く同じ一般リスクを持つ、つまり

$$\text{すべての}\theta\text{に対し、}R(\theta, d) = R(\theta, d')$$

となるものは、同じ決定方式と評価される。こうなるとそもそも比較すること自体必要がなくなる。従って、式 $(*)$ ですべてが等しい場合を除く。あらゆる $\theta$ ついて見渡して、

$$\text{すべての}\theta\text{に対し、}R(\theta,d) \leq R(\theta,d'),$$
$$\text{ある}\theta\text{に対し、}R(\theta,d) < R(\theta,d')$$

ならば、このことを簡単に

$$d \succ d'$$

と表し、「$d$ は $d'$ に優越する」といい、$\succ$ を自然な順序という。

一般に、決定方式 $d$ に対して、

$$d' \succ d \text{ となる } d' \text{は存在しない}$$

ならば、$d$ は許容的であるという。これは"一応は候補として残しておく"ということを意味する。この否定は、つまり、決定方式 $d$ に対して

$$d' \succ d \text{ となる } d' \text{がある}$$

ならば、$d$ は非許容的であるという。言い換えると、この時点で"消してよい"決定方式である。許容的ということは、積極的な良さとは異なり、それより良い（$\succ$ の意味で）ものはない（しかし最良ではない）という消極的な良さにとどまる。それに反して、非許容的な方式を用いるということは、それより良いものをさしおいて、それを用いる、という明白にまずい決定をしているのである。

例の場合、$d_1$と $d_2$〜$d_9$ における一般リスクの比較は、

$$\begin{array}{lll} R(\theta_1,d_1) < R(\theta_1,d_2), & R(\theta_1,d_1) < R(\theta_1,d_3), & R(\theta_1,d_1) < R(\theta_1,d_4) \\ R(\theta_2,d_1) > R(\theta_2,d_2), & R(\theta_2,d_1) > R(\theta_2,d_3), & R(\theta_2,d_1) > R(\theta_2,d_4) \\ & & \\ R(\theta_1,d_1) < R(\theta_1,d_5), & R(\theta_1,d_1) < R(\theta_1,d_6), & R(\theta_1,d_1) < R(\theta_1,d_7) \\ R(\theta_2,d_1) > R(\theta_2,d_5), & R(\theta_2,d_1) > R(\theta_2,d_6), & R(\theta_2,d_1) > R(\theta_2,d_7) \\ & & \\ R(\theta_1,d_1) < R(\theta_1,d_8), & R(\theta_1,d_1) < R(\theta_1,d_9) & \\ R(\theta_2,d_1) > R(\theta_2,d_8), & R(\theta_2,d_1) > R(\theta_2,d_9) & \end{array}$$

である。これを残りの一般リスクも同様に比較すると資料 3 になる。

これより、$d_2 \succ d_4, d_2 \succ d_7, d_3 \succ d_4, d_3 \succ d_5, d_3 \succ d_7, d_3 \succ d_8, d_5 \succ d_7, d_6 \succ d_7, d_6 \succ d_8$ となるため、$d_1, d_2, d_3, d_6, d_9$は許容的であり、$d_4, d_5, d_7, d_8$ は非許容的であることがわかる。また、決定方式 $d_3$ に優越されるような決定方式 $d$ は、図 1 のより、$d_4, d_5, d_7, d_8$ であることがわかる。また、$d_1, d_2, d_3, d_6, d_9$ の間に優越関係が存在

しないことも知ることができる。このように、状態が $\theta_1, \theta_2$ の 2 つの状態の場合には、リスク集合を平面図上に表すことで、決定方式間の優越関係を簡単に知ることができる。

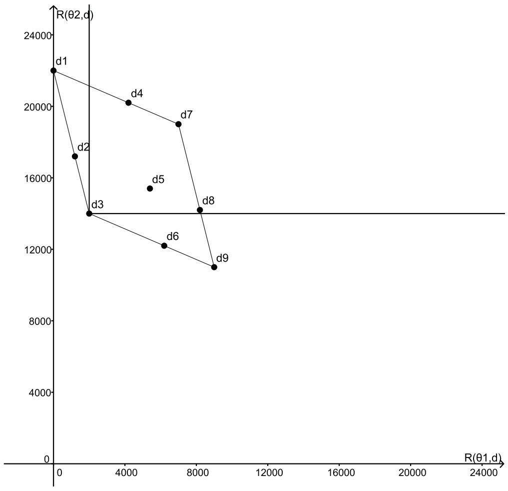


図 1: 例のリスク点とリスク集合

## 第 3 節　ミニマックス決定方式とベイズ決定方式

一般リスクは、決定方式を選別するのに役立った。特に、自然な順序 $\succ$ が決定方式どうしを比較するために導入され、それを用いて許容性という一応の選別基準が考えられることになった。しかし、何らかの理由でどれか 1 つを選ぶ場合になると、一般リスクよりもう一歩踏み込んだ概念を用いなければ不可能である。ミニマックス決定方式、ベイズ決定方式は、このような目的に役に立つ。

### ミニマックス決定方式

ミニマックス決定方式は、ゲーム理論の正統的手続きに従い、一般リスクを最大限に見積もった量

$$\max\{R(\theta_1, d), R(\theta_2, d), ..., R(\theta_I, d)\}$$

が最小の $d$ を用いる、という考えに基づく。ミニマックス原理は、ゲーム理論でなくても、合理的発想によるなら自然に思いつく直観的な判断方式である。

例の場合、以下のような量で比べる。

| | $d_1$ | $d_2$ | $d_3$ | $d_6$ | $d_9$ |
|---|---|---|---|---|---|
| リスクを最大限見積もった量 | 22000 | 17200 | 14000 | 12200 | 11000 |

表 16: 一般リスクを最大限見積もった量

表 16 から、ミニマックス決定方式は $d_9$ となる。従って、ミニマックス決定方式 $d_9$ により、$z_1$ が観測されたときには、行動 $a_3$ をとり、$z_2$ が観測されたときにも、行動 $a_3$ をとることになる。

平面図上で見るために、

$$\max\{R(\theta_1, d), R(\theta_2, d)\} = c \qquad (c > 0)$$

を満足する点を考えると、横軸 $c$ の位置から上方に立てた長さ $c$ の垂直な線分と縦軸 $c$ の位置から右方に立てた長さ $c$ の横軸に平行な線分を合わせたもの、逆 L 型群になる。従って、$c$ を徐々に大きくしていったときに、逆 L 型群とはじめて接するリスク点がミニマックス決定方式となる。

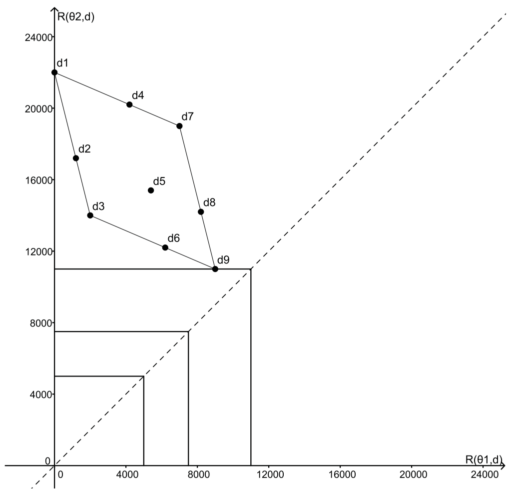


図 2: ミニマックス決定方式の図解

**ベイズ決定方式**

ベイズ決定方式は、自然の状態 $\theta_1, \theta_2, ..., \theta_I$ の事前確率 $w_1, w_2, ..., w_I$ が使えるならば、各リスク $R(\theta_1, d), R(\theta_2, d), ..., R(\theta_I, d)$ もこの確率で出現すると考えられることから、その期待値（期待リスク）

$$w_1R(\theta_1, d) + w_2R(\theta_2, d) + ... + w_IR(\theta_I, d)$$

が最小の $d$ を用いる、という考えに基づく。

例の場合、事前確率は

$$w_1 = \frac{7}{10}, \quad w_2 = \frac{3}{10}$$

となるため、許容的な決定方式の各期待リスクは、

$$\begin{aligned} w_1R(\theta_1, d_1) + w_2R(\theta_2, d_1) &= \frac{7}{10} \times 0 + \frac{3}{10} \times 22000 \\ &= 6600 \end{aligned}$$

$$
\begin{aligned}
w_1R(\theta_1,d_2)+w_2R(\theta_2,d_2) &= \frac{7}{10}\times 1200+\frac{3}{10}\times 17200 \\
&= 6000
\end{aligned}
$$

$$
\begin{aligned}
w_1R(\theta_1,d_3)+w_2R(\theta_2,d_3) &= \frac{7}{10}\times 2000+\frac{3}{10}\times 14000 \\
&= 5600
\end{aligned}
$$

$$
\begin{aligned}
w_1R(\theta_1,d_6)+w_2R(\theta_2,d_6) &= \frac{7}{10}\times 6200+\frac{3}{10}\times 12200 \\
&= 8000
\end{aligned}
$$

$$
\begin{aligned}
w_1R(\theta_1,d_9)+w_2R(\theta_2,d_9) &= \frac{7}{10}\times 9000+\frac{3}{10}\times 11000 \\
&= 9600
\end{aligned}
$$

| | $d_1$ | $d_2$ | $d_3$ | $d_6$ | $d_9$ |
|---|---|---|---|---|---|
| 期待リスク | 6600 | 6000 | 5600 | 8000 | 9600 |

表 17: 期待リスク

となり、期待リスクが最小となるのは、$d_3$ である。従って、ベイズ決定方式 $d_3$ により、$z_1$ が観測されたときには、行動 $a_1$ をとり、$z_2$ が観測されたときには、行動 $a_3$ をとることになる。

平面図上で見るために、

$$R(\theta_1,d)=x, R(\theta_2,d)=y$$

とおけば、期待リスク $r$ は、次のようになる。

$$\frac{7}{10}x+\frac{3}{10}y=r \qquad (r>0)$$

右辺の値を増やすとともに右上方に移動し、この直線とはじめて接するリスク点がベイズ決定方式となる。

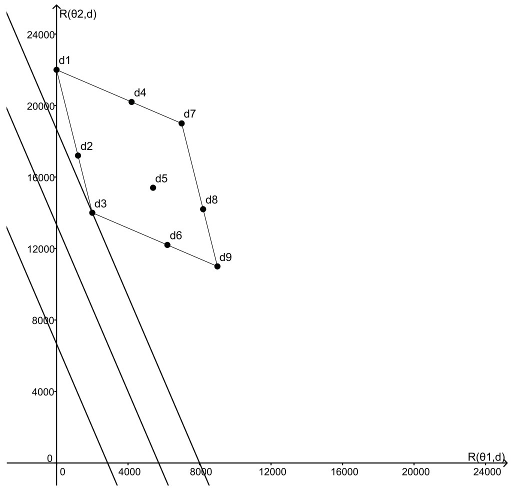


図 3: ベイズ決定方式の図解

ところで、ある事前確率に対するベイズ決定方式のみを知りたいというような場合には、これまでのように、すべての決定方式の一般リスクを求め、次に事前確率を用いて期待リスクを求めるような手順を踏まなくても、ベイズの定理を用いて、事後確率と損失関数を利用した期待値によってベイズ決定方式を求めることができる。以上のことを事後期待損失最小化の定理といい、証明は以下である。

定理（事後期待損失最小化）

各 $z$ に対し

$$\min_a \sum_\theta w'(\theta|z)L(\theta, a)$$

を解き、その解を $a^* = d(z)$ とおけば、$d$ は $R(\theta, d)$ を最小化する。

（証明）

各データ $z$ をそれに最適な行動 $a$ に写像する関数を

$$a = d(z)$$

とおくと、$d$ を用いるときのリスクは

$$R(\theta, d) = \sum_z p(z|\theta)L(\theta, d(z))$$

となる。これは $\theta$ によって変わる量であるから、$\theta$ についてさらに平均すると、和

の変更を経て

$$\begin{aligned} r(d) &= \sum_{\theta} w'(\theta)R(\theta,d) = \sum_{\theta}\sum_{z} w'(\theta)p(z|\theta)L(\theta,d(z)) \\ &= \sum_{z} p(z)\{\sum_{\theta} w'(\theta|z)L(\theta,d(z))\} \end{aligned}$$

を得る。ここに $p(z)$ はベイズの定理の分母である。

ここで、$r(d)$ を最小にする $d$ を”最適”ということにし、それを「ベイズ決定」といえば、まさに事後期待損失 $\{\}$ の最小化することに等しい。

この事後期待損失の定理によって、第 2 章第 3 節、第 4 節におけるリスクが最小であることを保証している。

# 終章　総括と今後の課題

## 第1節　総括

本研究の目的は、

- ベイズ理論が、意思決定において、頻度主義を基にした確率・統計学に比べ、適応できる場合が多い理論であることを示すこと
- 高等学校の確率・統計教育への導入が可能であることを示すこと

であった。

まず、各章の総括を述べる。

第1章；ベイズ理論の歴史から、ワルド、シュレイファー、ライファという意思決定にベイズ理論を利用した人たちは、全員、初めからベイズ理論を支持していたのではなく、頻度主義を基にした統計学を学んでいたり、教えたりしていた。それにもかかわらず、意思決定に頻度主義の統計学ではなく、ベイズ理論を重視していた。

第2章；ベイズ理論の基礎となるベイズの定理を確認、証明するとともに、ベイズの定理によって、データが得られる度に事後確率が更新されていく、ベイズ更新を導いた。また、意思決定におけるリスクの考え方を定義し、具体例を2つ取り上げた。

第3章；決定理論を抽象化し、その中で代表的な決定方式であるミニマックス決定方式とベイズ決定方式を定義し、状態が2つの場合の具体例に用いた。また、ベイズ決定方式から導くことのできる事後期待損失最小化の定理を証明した。この事後期待損失最小化の定理が第2章で述べたリスクの最小化を保証することがわかった。

以上から、

- ベイズ理論が頻度主義を基にした確率・統計学に比べ、意思決定において適応できる場合が多い。その理由として、頻度主義では同一条件で無限に繰り返す試行が前提とされ、確率を1つの値に定めるのに対し、ベイズ理論では事前確率を設定し、その事前確率と少なくとも1つのデータと組み合わせることで事後確率を求める、そして、その事後確率とデータを組み合わせるベイズ更新をすることよって、確率を1つの値に定めることにある

- 高等学校における確率・統計教育への導入が可能である。その理由は、第２章、第３章の内容が数学Ａの（場合の数と確率）と数学Ｂの（確率変数と確率分布）を既習していれば、指導可能と判断できるからである

と結論できる。

従って、意思決定におけるベイズ理論の有用性を学ぶことで、本研究の目的が達成できることが分かった。

## 第２節　今後の課題

本研究の今後の課題としては、実践における課題２点を挙げる。

(1) 時間の関係上、残念ながら、本研究について実際に授業を行い、考察し、成果と課題を明らかにすることができなかった。
そのため、ここでは実際に行う際の指導と評価の計画を提示する。

| 時 | 学習内容 | 評価基準 | 評価方法 |
|---|---|---|---|
| 1,2 | 条件付き確率、乗法定理を用いて、ベイズの定理、ベイズ更新を求める。 | ベイズの定理やベイズ更新に関心を持ち、具体的な事象の考察に活用しようとする。(関心・意欲・態度)<br>具体的な問題に、ベイズの定理やベイズ更新を用いることで、原因の確率、確率が更新されることを考察することができる。(数学的な見方・考え方) | 観察、課題、プリント |
| 3 | リスクについて理解し、ベイズ理論を用いて、リスク最小にする意思決定を求める。 | ベイズの定理、ベイズ更新を使って、事後確率を求めることができる。<br>(数学的な技能)<br>リスクについて理解している。<br>(知識・理解) | 観察、課題、プリント |

この指導と評価の計画（３時間扱い）は第２章の部分となっており、既習事項として確率の基本的な性質、余事象、独立、条件付き確率、乗法定理、平均（期待値）を必要とするため、数学Ａの（場合の数と確率)、数学Ｂの（確率変数と確率分布）を終えている高等学校２学年であれば行うことのできる授業計画となっている。また、資料４に学習指導案を加えている。

| 時 | 学習内容 | 評価基準 | 評価方法 |
|---|---|---|---|
| 1,2 | 決定理論を抽象化し、意思決定方式であるミニマックス決定方式とベイズ決定方式を理解する。事後期待損失最小化の定理を証明する。 | 決定理論に関心を持ち、ミニマックス決定方式、ベイズ決定方式を具体的な事象の考察に活用しようとする。<br>(関心・意欲・態度)<br>事後期待損失最小化の定理が、ベイズ理論におけるリスクが最小であることを保証する、ということを理解している。(知識・理解) | 観察、課題、プリント |

この指導と評価の計画（2 時間扱い）は第 3 章の部分となっており、第 2 章で示している部分を終えていることが必要となる。この授業を行う対象としては、高等学校 2 学年であれば行うことのできる授業計画となっている。また、資料 3 に学習指導案を加えている。

(2) 実際に授業を行う際には、第 2 章、第 3 章で取り上げた具体例だけでなく、高校生が身近に感じることのできる教材を探し、取り上げることが重要である。従って、そこには十分に時間をかけて考える必要がある。

# 付章　ベイズ統計学とカルバック・ライブラーの情報量

付章第1節では、ベイズの定理を確率の形から統計学の形へと発展させる。

付章第2節では、情報理論におけるカルバック・ライブラーの情報量を定義し、性質について証明する。

付章第3節では、カルバック・ライブラーの情報量を用いることで、ベイズ統計学が頻度主義を基にした統計学と同様に扱えることを明らかにする。

## 第1節　ベイズ統計学

ベイズの定理で、原因 $(B)$ を$\theta$、結果 $(A)$ を $z$、事前確率 $P(B_i)$ を $w(\theta_i)$、事後確率 $P(Bi|A)$ を $w'(\theta_i|z)$、確率 $P(A|B_i)$ を $z$ の関数として $p(z|\theta_i)$ と書けば、
ベイズの定理は、

$$\text{離散型}\quad w'(\theta_i|z) = \frac{w(\theta_i)p(z|\theta_i)}{\sum_j w(\theta_j)p(z|\theta_j)}$$

となる。
さらに、$\theta$が連続的に動くならば、上式は積分を用いて

$$\text{連続型}\quad w'(\theta|z) = \frac{w(\theta)p(z|\theta)}{\int_\Theta w(\theta)p(z|\theta)d\theta}$$

の形となる。ここで、$\Theta$は$\theta$全体の集合である。
2 式の統計学的な解釈として、
$\theta$ : 母集団のパラメータ（母数）
$z$ : サンプル
$p(z|\theta)$ : パラメータが$\theta$の時のサンプルの確率分布、言い換えると、$z$ を得たときの$\theta$の尤度
$w(\theta)$ : 事前分布
$w'(\theta|z)$ : 事後分布
となる。ここで、2 式の分母について考えると、

$$\int_\Theta w(\theta)p(z|\theta)d\theta = p(z)$$

となる。さて、データが得られた段階では、具体的な値はともかくとして $p(z)$ は定数になるはずである。それは、すでにデータは得られ確定しているからである。そこで、$p(z)$ を定数と考え、$k = \frac{1}{p(z)}$ とすると、

$$w'(\theta|z) = kw(\theta)p(z|\theta)$$

となる。
さらに、計算されないことが多く考慮しないため $k$ の部分を省略して、

$$w'(\theta|z) \propto w(\theta)p(z|\theta)$$

となる。
これをわかりやすく言えば、

$$\text{事後に有する情報} = \text{事前に関する情報} + \text{サンプルの出方の情報}$$

ということである。

予測分布
今、2 つの確率変数 $y$、$y'$があり、順番として $y$ の後に $y'$が続くものと考える。2 つともパラメータ$\theta$によって決まる分布を持つとすれば、$y \Rightarrow \theta \Rightarrow y'$の因果序列で、$\theta$の事後分布を媒介して、$y$ の分布から $y'$の分布が求められる。一般にこの考え方を予測分布という。$f$ を $y, y'$などの各分布を表す一般記号とし、$w'$で$\theta$の分布を表す。
同時分布としては、

$$f(y', y) = \int f(y', y, \theta)d\theta$$

であり、この上式の両辺を $y$ の周辺分布 $f(y)$ で割ると、

$$\frac{f(y', y)}{f(y)} = \int \frac{f(y', y, \theta)}{f(y)}d\theta$$

となる。この右辺のインテグラルの中を変形すると、

$$\begin{aligned} \frac{f(y', y, \theta)}{f(y)} &= \frac{f(y'|\theta, y)f(\theta, y)}{f(y)} \\ &= f(y'|\theta, y)w'(\theta|y) \end{aligned}$$

であり、左辺は、

$$\frac{f(y', y)}{f(y)} = f(y'|y)$$

となるから、従って、

$$f(y'|y) = \int f(y'|\theta, y)w'(\theta|y)d\theta$$

が得られる。多くの場合、$y'$の分布は$\theta$だけの関数として設定され、

$$f(y'|y) = \int f(y'|\theta)w'(\theta|y)d\theta$$

の形に帰する。

## 第2節　カルバック・ライブラーの情報量

定義

$\boldsymbol{q}$、$\boldsymbol{q'}$を2つの確率分布を表すベクトルとする。

$$\boldsymbol{q} = (q_1, q_2, \cdots, q_l), \boldsymbol{q'} = (q'_1, q'_2, \cdots, q'_l)$$

ここで、

$$q_1 + q_2 + \cdots + q_l = q'_1 + q'_2 + \cdots + q'_l = 1$$

$$q_1 > 0, q_2 > 0, \cdots, q_l > 0, \quad q'_1 > 0, q'_2 > 0, \cdots, q'_l > 0$$

このとき、カルバック・ライブラーの情報量と言われる量を

$$I(\boldsymbol{q}、\boldsymbol{q'}) = q_1 \log \frac{q_1}{q'_1} + q_2 \log \frac{q_2}{q'_2} + \cdots + q_l \log \frac{q_l}{q'_l}$$

で定義する。

このとき、次の性質が成り立つ。

$(a) I(\boldsymbol{q}、\boldsymbol{q'}) \geq 0$

$(b) \boldsymbol{q} = \boldsymbol{q'}$ならば $I(\boldsymbol{q}、\boldsymbol{q'}) = 0$。逆も真である。

**証明 (a)**

正数 $u$ の関数

$$\phi(u) = u - 1 - \log u$$

を考えると、

$$\begin{aligned}
&\phi(1) = 1 - 1 - \log 1 = 0 \\
&\phi'(u) = 1 - (\frac{1}{u}) \\
&\phi'(1) = 1 - (\frac{1}{1}) = 0 \\
&\phi'(u) = 1 - (\frac{1}{u}) > 0 (u > 1) \\
&\phi'(u) < 0 (0 < u < 1)
\end{aligned}$$

であるから、$u = 1$ で最小値 0 をとる。従って、$\log u \leq u - 1$ である。

故に、$x, y$ を正の数とすると

$$\begin{aligned}
\log \frac{y}{x} &\leq \frac{y}{x} - 1 \\
x \log \frac{y}{x} &\leq y - x \\
x \log \frac{x}{y} &\geq x - y
\end{aligned}$$

ここで、$x = q_1, y = q'_1$とすると

$$q_1 \log \frac{q_1}{q'_1} \geq q_1 - q'_1$$

$x = q_2, y = q'_2$のとき

$$q_2 \log \frac{q_2}{q'_2} \geq q_2 - q'_2$$

これを $x = q_l, y = q'_l$まで代入すると

$$q_1 \log \frac{q_1}{q'_1} \geq q_1 - q'_1, q_2 \log \frac{q_2}{q'_2} \geq q_2 - q'_2, \cdots, q_l \log \frac{q_l}{q'_l} \geq q_l - q'_l$$

この $l$ 個の不等式を加え合わせると、

$$\begin{aligned} q_1 \log \frac{q_1}{q'_1} + q_2 \log \frac{q_2}{q'_2} + \cdots + q_l \log \frac{q_l}{q'_l} &\geq q_1 + q_2 + \cdots + q_l - (q'_1 + q'_2 + \cdots + q'_l) \\ &= 0 \end{aligned}$$

**証明 (b)**

$\boldsymbol{q} = \boldsymbol{q'}$のとき $(\Rightarrow)$

$$\begin{aligned} q_1 \log 1 + q_2 \log 1 + \cdots + q_l \log 1 &= 0 + 0 + \cdots + 0 \\ &= 0 \\ I(\boldsymbol{q}、\boldsymbol{q'}) &= 0 \end{aligned}$$

$I(\boldsymbol{q}、\boldsymbol{q'}) = 0$ のとき $(\Leftarrow)$

$$\begin{aligned} \sum_{i=1}^{l} q_i \log \frac{q_i}{q'_i} &= \sum_{i=1}^{l} q_i - \sum_{i=1}^{l} q'_i \\ &= \sum_{i=1}^{l} q_i (1 - \frac{q'_i}{q_i}) \\ \log \frac{q_i}{q'_i} &= 1 - \frac{q'_i}{q_i} \\ \log \frac{q'_i}{q_i} &= \frac{q'_i}{q_i} - 1 \\ \log u &= u - 1 \\ \text{これは、} u &= 1 \text{ なので} \\ \frac{q'_i}{q_i} = 1, & \quad q'_i = q_i \end{aligned}$$

また、密度関数 $f(x), g(x)$ のときも

$$I(f, g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx$$

と定義すれば、$(a), (b)$ を満足する。

## 第 3 節　ベイズ統計学とカルバック・ライブラー情報量

ベイズ統計学では、データ $x$ が与えられると、事前分布 $p(\theta)$ は、事後分布 $p(\theta|x) = \frac{p(\theta)f(x|\theta)}{f(x)}$ に置き換える。ここに $f(x) = \int p(\theta)f(x|\theta)d\theta$ である。なぜこのような置き換えがなされなければならないのであろうか。

ベイズ統計学の目的を将来の観測値の分布の有効な決定と見ることによって、この疑問に答えることができる。将来の観測値 $y$ が、データ $x$ と同一の分布 $f(\cdot|\theta)$ に従って独立に分布するものとする。

この時の予測分布 $p(y|x)$ について考える。

カルバック・ライブラーの情報量より

$$\begin{aligned} I(f(y|\theta), p(y|x)) &= \int f(y|\theta)\log\frac{f(y|\theta)}{p(y|x)}dy \\ &= \int f(y|\theta)\log f(y|\theta)dy - \int f(y|\theta)\log p(y|x)dy \end{aligned}$$

予測分布の良さ（予測分布がどれだけ真の分布が近いか）というのは、上式の第 2 項で評価する。

ここで、予測分布の平均的な良さは

$$\iint f(x|\theta)f(y|\theta)\log p(y|x)dydx$$

となる。$f(x) = \int p(\theta)f(x|\theta)d\theta$ とすれば、上の積分は

$$\iiint f(y|\theta)\{\frac{p(\theta)f(x|\theta)}{f(x)}\}d\theta\log p(y|x)dyf(x)dx$$

と書ける。

$\int f(y|\theta)\{\frac{p(\theta)f(x|\theta)}{f(x)}\}d\theta$ が $y$ に関する一つの確率分布となっていることに注意すると、この分布に関する平均対数尤度が最大となるのは、$p(y|x)$ がこの分布に一致するときである。

従って、この場合最良の予測分布は

$$p(y|x) = \int f(y|\theta)p(\theta|x)d\theta$$

によって与えられる。ただし、$p(\theta|x) = \frac{p(\theta)f(x|\theta)}{f(x)}$ である。

こうして、事後分布 $p(\theta|x)$ は、平均的な意味で最良の予測分布を与える分布であることがわかる。

一般に、$p(\theta|x)$ を $x$ の関数として与えられる $\theta$ の任意の分布を考え、これを推測分布と呼ぶことにする。推測分布の役割は、$p(y|x) = \int f(y|\theta)p(\theta|x)d\theta$ によって予

測分布を定義することである。

今、$x$ と $y$ が独立に同一の分布 $f(\cdot)$ に従うものとし、$p(y|x)$ を用いて $f(y)$ を推定することが考える。

このとき、カルバック・ライブラーの情報量より

$$\begin{aligned}
I(f(y), p(y|x)) &= \int f(y) \log \frac{f(x)}{p(y|x)} dy \\
&= \int f(y) \log f(y) dy - \int f(y) \log p(y|x) dy
\end{aligned}$$

上式の $f(x)$ に関する平均は

$$\begin{aligned}
\int f(x)\{I(f(y), p(y|x))\} &= \int f(x) \int f(y) \log f(y) dy dx - \int f(x) \int f(y) \log p(y|x) dy dx \\
&= \int f(y) \log f(y) dy - \iint f(x) f(y) \log p(y|x) dy dx \qquad (9)
\end{aligned}$$

ここで（9）を次のように考える。

$$\begin{aligned}
(9) &= \int f(y) \log f(y) dy - \int f(y) \log\{\int f(x) p(y|x) dx\} dy \\
&+ \int f(y) \log\{\int f(x) p(y|x) dx\} dy - \iint f(x) f(y) \log p(y|x) dy dx
\end{aligned}$$

$$\begin{aligned}
B\{p(\theta|x)\} &= \int f(y) \log f(y) dy - \int f(y) \log\{\int f(x) p(y|x) dx\} dy \\
V\{p(\theta|x)\} &= \int f(y) \log\{\int f(x) p(y|x) dx\} dy - \iint f(x) f(y) \log p(y|x) dy dx
\end{aligned}$$

とすると、

$$(14) = B\{p(\theta|x)\} + V\{p(\theta|x)\}$$

となる。また、

$$\begin{aligned}
B\{p(\theta|x)\} &= \int f(y) \log f(y) dy - \int f(y) \log\{\int f(x) p(y|x) dx\} dy \\
&= \int f(y) \log f(y) dy - \int f(y) \log\{\int f(x) \int f(y|\theta) p(\theta|x) d\theta dx\} dy \\
&= \int f(y) \log f(y) dy - \int f(y) \log\{\int f(y|\theta) \int f(x) p(\theta|x) dx d\theta\} dy
\end{aligned}$$

$$\begin{aligned}
V\{p(\theta|x)\} &= \int f(y) \log\{\int f(x) p(y|x) dx\} dy - \iint f(x) f(y) \log p(y|x) dy dx \\
&= \int f(y) \log\{\int f(y|\theta) \int f(x) p(\theta|x) dx d\theta\} dy \\
&\qquad - \iint f(y) f(x) \log\{\int f(y|\theta) p(\theta|x) d\theta\} dx dy
\end{aligned}$$

$B\{p(\theta|x)\}$、$V\{p(\theta|x)\}$ はそれぞれ負となることはなく、
$B\{p(\theta|x)\}$ は、$\int f(x)p(\theta|x)dx$ によって定義される推測分布に基づく予測分布の $f(y)$ からの隔たりを与える。これは、統計的推定量の偏りに相当する量である。
$V\{p(\theta|x)\}$ は、$p(\theta|x)$ が $\int f(x)p(\theta|x)dx$ のまわりに変動するために生じる予測分布の平均的な劣化の程度を測るもので、通常の推定量の分散に相当する。
$x$ に依存する $p(\theta|x)$ の変動の範囲を小さく押さえれば、$V\{p(\theta|x)\}$ は小にできるが、$B\{p(\theta|x)\}$ が小さいとは言えない。また、$p(\theta|x)$ が良く $x$ の動きに反応すれば、$B\{p(\theta|x)\}$ は小にできるが、$V\{p(\theta|x)\}$ が小さいとは言えない。

統計的推論の問題は、この $B\{p(\theta|x)\}$ と $V\{p(\theta|x)\}$ との間の適当なバランスを実現するような $p(\theta|x)$ を見出すことによって、$B\{p(\theta|x)\} + V\{p(\theta|x)\}$ を予想される $f(\cdot)$ の様々な場合に対して、小さく押さえることにある。

はじめの部分で最良の $p(y|x)$ を与えるのは、$\theta$ の事後分布であるという結果が得られた。また、一般に推測分布 $p(\theta|x)$ は任意に考えることができ、ベイズ統計学の事後分布はその 1 つである。

従って、エントロピーと予測分布の考えとを結合することによって、ベイズ統計学による事後分布が最良の形であり、統計学において抵抗なく扱えるという視点が得られた。この立場から見ると、統計的モデルの構成あるいは選択に際しては、平均的な意味でエントロピーを大きくするように心がけるべきであることがわかる。

# 資料

## 資料１　第２章の例１における事後確率とリスク

事前確率を

$$P(A|B_1^c) = 0.467, \quad P(A^c|B_1^c) = 0.533$$

とする。
ここで、ベイズの定理を用いて、事後確率を算出すると、

$$\begin{aligned} P(A|B_1^c \cap B_2) &= \frac{P(A|B_1^c)P(B_2|A)}{P(A|B_1^c)P(B_2|A) + P(A^c|B_1^c)P(B_2|A^c)} \\ &= \frac{0.467 \times (\frac{7}{10})}{0.467 \times (\frac{7}{10}) + 0.533 \times (\frac{2}{10})} = 0.754 \end{aligned}$$

$$\begin{aligned} P(A|B_1^c \cap B_2^c) &= \frac{P(A|B_1^c)P(B_2^c|A)}{P(A|B_1^c)P(B_2^c|A) + P(A^c|B_1^c)P(B_2^c|A^c)} \\ &= \frac{0.467 \times (\frac{3}{10})}{0.467 \times (\frac{3}{10}) + 0.533 \times (\frac{8}{10})} = 0.247 \end{aligned}$$

$$\begin{aligned} P(A^c|B_1^c \cap B_2) &= \frac{P(A^c|B_1^c)P(B_2|A^c)}{P(A|B_1^c)P(B_2|A) + P(A^c|B_1^c)P(B_2|A^c)} \\ &= \frac{0.533 \times (\frac{2}{10})}{0.467 \times (\frac{7}{10}) + 0.533 \times (\frac{2}{10})} = 0.246 \end{aligned}$$

$$\begin{aligned} P(A^c|B_1^c \cap B_2^c) &= \frac{P(A^c|B_1^c)P(B_2^c|A^c)}{P(A|B_1^c)P(B_2^c|A) + P(A^c|B_1^c)P(B_2^c|A^c)} \\ &= \frac{0.533 \times (\frac{8}{10})}{0.467 \times (\frac{3}{10}) + 0.533 \times (\frac{8}{10})} = 0.753 \end{aligned}$$

である。

午前の天気予報が「晴れまたは曇り」、午後の天気予報は「雨」と発表したときのリスクは

$$R(B_1 \cap B_2^c, a_1) = 0.754 \times 0 + 0.246 \times 20000 = 4920$$
$$R(B_1 \cap B_2^c, a_2) = 0.754 \times 6000 + 0.246 \times 14000 = 7968$$
$$R(B_1 \cap B_2^c, a_3) = 0.754 \times 10000 + 0.246 \times 10000 = 10000$$

となる。

午前の天気予報が「雨」、午後の天気予報は「晴れまたは曇り」と発表したときのリスクは

$$
\begin{aligned}
R(B_1^c \cap B_2, a_1) &= 0.754 \times 0 + 0.246 \times 20000 = 4920 \\
R(B_1^c \cap B_2, a_2) &= 0.754 \times 6000 + 0.246 \times 14000 = 7968 \\
R(B_1^c \cap B_2, a_3) &= 0.754 \times 10000 + 0.246 \times 10000 = 10000
\end{aligned}
$$

となる。

午前の天気予報が「雨」、午後の天気予報は「雨」と発表したときのリスクは

$$
\begin{aligned}
R(B_1^c \cap B_2^c, a_1) &= 0.247 \times 0 + 0.753 \times 20000 = 15060 \\
R(B_1^c \cap B_2^c, a_2) &= 0.247 \times 6000 + 0.753 \times 14000 = 12024 \\
R(B_1^c \cap B_2^c, a_3) &= 0.247 \times 10000 + 0.753 \times 10000 = 10000
\end{aligned}
$$

となる。

## 資料２　第３章の例における各一般リスク

$$\begin{aligned}
R(\theta_1, d_2) &= p(z_1|\theta_1)L(\theta_1, d_2(z_1)) + p(z_2|\theta_1)L(\theta_1, d_2(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_1) + p(z_2|\theta_1)L(\theta_1, a_2) \\
&= \frac{7}{10} \times 0 + \frac{2}{10} \times 6000 = 1200
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_2) &= p(z_1|\theta_2)L(\theta_2, d_2(z_1)) + p(z_2|\theta_2)L(\theta_2, d_2(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_1) + p(z_2|\theta_2)L(\theta_2, a_2) \\
&= \frac{3}{10} \times 20000 + \frac{8}{10} \times 14000 = 17200
\end{aligned}$$

$$\begin{aligned}
R(\theta_1, d_3) &= p(z_1|\theta_1)L(\theta_1, d_3(z_1)) + p(z_2|\theta_1)L(\theta_1, d_3(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_1) + p(z_2|\theta_1)L(\theta_1, a_3) \\
&= \frac{7}{10} \times 0 + \frac{2}{10} \times 10000 = 2000
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_3) &= p(z_1|\theta_2)L(\theta_2, d_3(z_1)) + p(z_2|\theta_2)L(\theta_2, d_3(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_1) + p(z_2|\theta_2)L(\theta_2, a_3) \\
&= \frac{3}{10} \times 20000 + \frac{8}{10} \times 10000 = 14000
\end{aligned}$$

$$\begin{aligned}
R(\theta_1, d_4) &= p(z_1|\theta_1)L(\theta_1, d_4(z_1)) + p(z_2|\theta_1)L(\theta_1, d_4(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_2) + p(z_2|\theta_1)L(\theta_1, a_1) \\
&= \frac{7}{10} \times 6000 + \frac{2}{10} \times 0 = 4200
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_4) &= p(z_1|\theta_2)L(\theta_2, d_4(z_1)) + p(z_2|\theta_2)L(\theta_2, d_4(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_2) + p(z_2|\theta_2)L(\theta_2, a_1) \\
&= \frac{3}{10} \times 14000 + \frac{8}{10} \times 20000 = 20200
\end{aligned}$$

$$\begin{aligned}
R(\theta_1, d_5) &= p(z_1|\theta_1)L(\theta_1, d_5(z_1)) + p(z_2|\theta_1)L(\theta_1, d_5(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_2) + p(z_2|\theta_1)L(\theta_1, a_2) \\
&= \frac{7}{10} \times 6000 + \frac{2}{10} \times 6000 = 5400
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_5) &= p(z_1|\theta_2)L(\theta_2, d_5(z_1)) + p(z_2|\theta_2)L(\theta_2, d_5(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_2) + p(z_2|\theta_2)L(\theta_2, a_2) \\
&= \frac{3}{10} \times 14000 + \frac{8}{10} \times 14000 = 15400
\end{aligned}$$

$$\begin{aligned}
R(\theta_1, d_6) &= p(z_1|\theta_1)L(\theta_1, d_6(z_1)) + p(z_2|\theta_1)L(\theta_1, d_6(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_2) + p(z_2|\theta_1)L(\theta_1, a_3) \\
&= \frac{7}{10} \times 6000 + \frac{2}{10} \times 10000 = 6200
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_6) &= p(z_1|\theta_2)L(\theta_2, d_6(z_1)) + p(z_2|\theta_2)L(\theta_2, d_6(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_2) + p(z_2|\theta_2)L(\theta_2, a_3) \\
&= \frac{3}{10} \times 14000 + \frac{8}{10} \times 10000 = 12200
\end{aligned}$$

$$\begin{aligned}
R(\theta_1, d_7) &= p(z_1|\theta_1)L(\theta_1, d_7(z_1)) + p(z_2|\theta_1)L(\theta_1, d_7(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_3) + p(z_2|\theta_1)L(\theta_1, a_1) \\
&= \frac{7}{10} \times 10000 + \frac{2}{10} \times 0 = 7000
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_7) &= p(z_1|\theta_2)L(\theta_2, d_7(z_1)) + p(z_2|\theta_2)L(\theta_2, d_7(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_3) + p(z_2|\theta_2)L(\theta_2, a_1) \\
&= \frac{3}{10} \times 10000 + \frac{8}{10} \times 20000 = 19000
\end{aligned}$$

$$\begin{aligned}
R(\theta_1, d_8) &= p(z_1|\theta_1)L(\theta_1, d_8(z_1)) + p(z_2|\theta_1)L(\theta_1, d_8(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_3) + p(z_2|\theta_1)L(\theta_1, a_2) \\
&= \frac{7}{10} \times 10000 + \frac{2}{10} \times 6000 = 8200
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_8) &= p(z_1|\theta_2)L(\theta_2, d_8(z_1)) + p(z_2|\theta_2)L(\theta_2, d_8(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_3) + p(z_2|\theta_2)L(\theta_2, a_2) \\
&= \frac{3}{10} \times 10000 + \frac{8}{10} \times 14000 = 14200
\end{aligned}$$

$$\begin{aligned}
R(\theta_1, d_9) &= p(z_1|\theta_1)L(\theta_1, d_9(z_1)) + p(z_2|\theta_1)L(\theta_1, d_9(z_2)) \\
&= p(z_1|\theta_1)L(\theta_1, a_3) + p(z_2|\theta_1)L(\theta_1, a_3) \\
&= \frac{7}{10} \times 10000 + \frac{2}{10} \times 10000 = 9000
\end{aligned}$$

$$\begin{aligned}
R(\theta_2, d_9) &= p(z_1|\theta_2)L(\theta_2, d_9(z_1)) + p(z_2|\theta_2)L(\theta_2, d_9(z_2)) \\
&= p(z_1|\theta_2)L(\theta_2, a_3) + p(z_2|\theta_2)L(\theta_2, a_3) \\
&= \frac{3}{10} \times 10000 + \frac{8}{10} \times 10000 = 11000
\end{aligned}$$

## 資料3　第3章の例における一般リスクの比較

$$\begin{array}{lll}
R(\theta_1, d_2) < R(\theta_1, d_3), & \underline{R(\theta_1, d_2) < R(\theta_1, d_4),} & R(\theta_1, d_2) < R(\theta_1, d_5) \\
R(\theta_2, d_2) > R(\theta_2, d_3), & \underline{R(\theta_2, d_2) < R(\theta_2, d_4),} & R(\theta_2, d_2) > R(\theta_2, d_5)
\end{array}$$

$$\begin{array}{lll}
R(\theta_1, d_2) < R(\theta_1, d_6), & \underline{R(\theta_1, d_2) < R(\theta_1, d_7),} & R(\theta_1, d_2) < R(\theta_1, d_8) \\
R(\theta_2, d_2) > R(\theta_2, d_6), & \underline{R(\theta_2, d_2) < R(\theta_2, d_7),} & R(\theta_2, d_2) > R(\theta_2, d_8)
\end{array}$$

$$\begin{array}{l}
R(\theta_1, d_2) < R(\theta_1, d_9) \\
R(\theta_2, d_2) > R(\theta_2, d_9)
\end{array}$$

$$\begin{array}{lll}
\underline{R(\theta_1, d_3) < R(\theta_1, d_4),} & \underline{R(\theta_1, d_3) < R(\theta_1, d_5),} & R(\theta_1, d_3) < R(\theta_1, d_6) \\
\underline{R(\theta_2, d_3) < R(\theta_2, d_4),} & \underline{R(\theta_2, d_3) < R(\theta_2, d_5),} & R(\theta_2, d_3) > R(\theta_2, d_6)
\end{array}$$

$$\begin{array}{lll}
\underline{R(\theta_1, d_3) < R(\theta_1, d_7),} & \underline{R(\theta_1, d_3) < R(\theta_1, d_8),} & R(\theta_1, d_2) < R(\theta_1, d_9) \\
\underline{R(\theta_2, d_3) < R(\theta_2, d_7),} & \underline{R(\theta_2, d_3) < R(\theta_2, d_8),} & R(\theta_2, d_2) > R(\theta_2, d_9)
\end{array}$$

$$\begin{array}{lll}
R(\theta_1, d_4) < R(\theta_1, d_5), & R(\theta_1, d_4) < R(\theta_1, d_6), & R(\theta_1, d_4) < R(\theta_1, d_7) \\
R(\theta_2, d_4) > R(\theta_2, d_5), & R(\theta_2, d_4) > R(\theta_2, d_6), & R(\theta_2, d_4) > R(\theta_2, d_7)
\end{array}$$

$$\begin{array}{ll}
R(\theta_1, d_4) < R(\theta_1, d_8), & R(\theta_1, d_4) < R(\theta_1, d_9) \\
R(\theta_2, d_4) > R(\theta_2, d_8), & R(\theta_2, d_4) > R(\theta_2, d_9)
\end{array}$$

$$\begin{array}{lll}
R(\theta_1, d_5) < R(\theta_1, d_6), & \underline{R(\theta_1, d_5) < R(\theta_1, d_7),} & R(\theta_1, d_5) < R(\theta_1, d_8) \\
R(\theta_2, d_5) > R(\theta_2, d_6), & \underline{R(\theta_2, d_5) < R(\theta_2, d_7),} & R(\theta_2, d_5) > R(\theta_2, d_8)
\end{array}$$

$$\begin{array}{l}
R(\theta_1, d_5) < R(\theta_1, d_9) \\
R(\theta_2, d_5) > R(\theta_2, d_9)
\end{array}$$

$$\begin{array}{lll}
\underline{R(\theta_1, d_6) < R(\theta_1, d_7),} & \underline{R(\theta_1, d_6) < R(\theta_1, d_8),} & R(\theta_1, d_6) < R(\theta_1, d_9) \\
\underline{R(\theta_2, d_6) < R(\theta_2, d_7),} & \underline{R(\theta_2, d_6) < R(\theta_2, d_8),} & R(\theta_2, d_6) > R(\theta_2, d_9)
\end{array}$$

$$\begin{array}{ll}
R(\theta_1, d_7) < R(\theta_1, d_8), & R(\theta_1, d_7) < R(\theta_1, d_9) \\
R(\theta_2, d_7) > R(\theta_2, d_8), & R(\theta_2, d_7) > R(\theta_2, d_9)
\end{array}$$

$$\begin{array}{l}
R(\theta_1, d_8) < R(\theta_1, d_9) \\
R(\theta_2, d_8) > R(\theta_2, d_9)
\end{array}$$

# 資料４　学習指導案

１　単元名　　　　「ベイズ理論」

２　本時の指導（１／３）

（１）題材名　「ベイズの定理」

（２）ねらい　具体例における結果に対する原因の確率を求めることで、ベイズの定理を理解する。また、定理を証明する。

（３）展開

| 主な学習活動と予想される生徒の反応 | ・指導上の留意点 |
|---|---|
| １．条件付き確率、乗法定理について振り返りをする。<br>S：条件付き確率は$P_A(B)=\frac{P(A\cap B)}{P(A)}$の形。<br>S：乗法定理は$P(A\cap B)=P(A)P_A(B)$の形。 | ・どのような式で表すことができたのか確認する。 |
| ベイズの定理を理解し、定理を証明する。 | |
| ２．結果に対する原因の確率を求める具体例を解く。　（病気の検査の問題）<br>S：結果の確率ばかり注目していて、原因の確率を考えていなかった。 | ・２つの事象を結果と原因の２つに分けることができることを確認する。<br>・原因に対する結果の確率と結果に対する原因の確率の違いを意識する。 |
| ３．ベイズの定理を説明し、証明する。 | ・定理を証明するのに、条件付き確率と乗法の定理が用いられることを確認する。<br>・原因の事象に関する条件を意識する。<br>・原因の事象が２つの場合（$A,A^c$の場合）を考えてから、k 個の場合を考える。 |
| ベイズの定理を用いることで、結果に対する原因の確率を求めることができる。また、ベイズの定理は、<br>$P_B(A_i)=\frac{P(A_i)P_{A_i}(B)}{\sum_{j=1}^{k}P(A_j)P_{A_j}(B)}$<br>である。 | |
| ５．練習問題を解く。 | |

1　単元名　　　「ベイズ理論」
2　本時の指導（２／３）
（１）題材名　「ベイズ更新」
（２）ねらい　ベイズ更新を一般化し、理解する。
（３）展開

<table>
<tr><th>主な学習活動と予想される生徒の反応</th><th>・指導上の留意点</th></tr>
<tr><td>１．ベイズの定理について振り返りをする。<br>S：ベイズの定理によって、結果に対する原因の確率を求めることができる。</td><td>・条件付き確率との違いをきちんと意識させる。</td></tr>
<tr><td colspan="2">結果の事象が何度も生じる場合、ベイズの定理を用いるとどうなるのかを求める。</td></tr>
<tr><td>２．結果の事象Bを何度も生じる事象と考えるときに、$t$ 回目の事象を$B_t$と表記し、原因の事象 A についてベイズの定理を使う。まずは、1回目を考える。<br>S：前時に求めたベイズの定理にあてはめながら解く。<br><br>３．次に、2回目について考える。<br>S：1回目の事後確率を2回目の事前確率として、ベイズの定理を用いる。<br>S：結果(データ)の確率として、$P(A|B_1 \cap B_2)$を使う。<br><br>４．n回目について考える。<br>S：2回目を参考に、$n-1$回目の事後確率を事前確率として利用する。</td><td>・原因の事象は、$A, A^C$の2つであり、互いに排反で、かつすべての場合をつくし、$A \cup A^c = \Omega$とすることを忘れない。<br><br>・2回だけなく、何度のベイズ更新が行えることを確認する。<br><br>・ベイズ更新の考え方が、人間の心理に近いことを確認する。<br><br>・具体例を用いて、逐次合理性を確認する。</td></tr>
<tr><td colspan="2">ベイズの定理で得られた事後確率を新たな事前確率として利用し、再びベイズの定理を用いて、次の事後確率を求めるという処理を繰り返すことをベイズ更新と呼ぶ。</td></tr>
<tr><td>５．練習問題を解く。</td><td></td></tr>
</table>

1　単元名　　　「ベイズ理論」

2　本時の指導（３／３）

（１）題材名　「意思決定におけるベイズ理論の有用性」

（２）ねらい　リスクについて理解し、意思決定にベイズ理論が利用できることを具体例から実感する。

（３）展開

| 主な学習活動と予想される生徒の反応 | ・指導上の留意点 |
|---|---|
| １．前時の振り返りをする。<br>S：ベイズの定理によって、結果に対する原因の確率を求めることができる。<br>S：結果が何度も得られる場合に、ベイズ更新を行うことができる。 | ・条件付き確率との違いをきちんと意識させる。 |
| ２．リスクを定義する。 | ・リスクを定義するときに、損失関数を式で表すことができないため、具体的にどのようなものなのか説明する。<br>・確率分布における期待値(平均)の求め方と同じであることを確認する。 |

具体例を用いて、意思決定にベイズ理論が利用できることを実感する。

| 主な学習活動と予想される生徒の反応 | ・指導上の留意点 |
|---|---|
| ３．具体例を解く。<br>S：天気予報を聞いた後の事後確率は、事前確率より、確率が高くなっていることがわかる。<br>S：「晴れまたは曇り」と発表した場合、行動$a_1$、「雨」と発表した場合、行動$a_3$をとる。 | ・損失関数によって求める損失値が結果によって異なることを意識させる。<br>・ベイズの定理によって、確率が変化することを確かめるとともに、リスクを最小にする行動が最良の意思決定であることを確認する。<br>・意思決定に確率、リスクの両方が使えることを確認する。 |

リスクを導入することで、意思決定にベイズ理論を利用することができる。また、リスクを最小にする行動が最良の意思決定である。

| 主な学習活動と予想される生徒の反応 | ・指導上の留意点 |
|---|---|
| ４．練習問題を解く。 | |

１　単元名　　　「統計的決定問題」

２　本時の指導（１／２）

（１）題材名　「意思決定方式と許容性」

（２）ねらい　決定理論について具体例を用いながら、各決定方式の一般リスクを比較し、決定方式の許容性について理解する。

（３）展開

| 主な学習活動と予想される生徒の反応 | ・指導上の留意点 |
|---|---|
| １．決定理論を抽象化する。 | ・意思決定におけるベイズ理論で扱った具体例と比較しながら、抽象化していく。 |
| ２．一般リスクを定義する。<br>S：各決定方式における一般リスクを求める。 | ・具体例の状態が 2 つであるため、各決定方式の一般リスクが 2 つあることを確認する。 |
| 具体例を用いて、決定方式の許容性について理解する。 | |
| ３．決定方式の許容性について説明し、具体例を用いながら各決定方式の一般リスクを比較し、許容性を確かめる。 | ・「すべての状態に対し」という部分をきちんと確認する。<br>・1 つでも優越される決定方式がある場合には、非許容的になることを意識させる。 |
| ４．Geogebra を使って、リスク点とリスク集合を平面上に図示し、図から許容性について考える。<br>S：決定方式の許容性は、リスク点から横軸と縦軸に平行な直線で見るとわかりやすい。<br>S：許容的とわかった決定方式は、リスク集合の左側にある。 | ・Geogebra の使い方を説明する。<br>・平面図からも許容性がわかるということを確認する。 |
| 1 つでも優越される決定方式があれば、非許容的であり、意思決定においては、許容的な決定方式を選ぶ。また、平面図を用いることで、許容性がよりわかりやすくなる。 | |
| ５．練習問題を解く。 | |

１　単元名　　　「統計的決定問題」
２　本時の指導（２／２）
　（１）題材名　「ミニマックス決定方式とベイズ決定方式」
　（２）ねらい　ミニマックス決定方式とベイズ決定方式について具体例を用いながら理解し、ベイズ決定方式から得られる事後期待損失最小化の定理を証明する。
　（３）展開

<table>
<tr><th>主な学習活動と予想される生徒の反応</th><th>・指導上の留意点</th></tr>
<tr><td>１．前時の振り返りをする。</td><td>・許容的な決定方式は複数あるが、決定方式を１つ選ぶ場合には、何か違う基準を考えなければならないことを確認する。</td></tr>
<tr><td colspan="2">ミニマックス決定方式とベイズ決定方式について理解し、事後期待損失最小化を証明する。</td></tr>
<tr><td>２．ミニマックス決定方式を説明し、具体例で確認する。<br>S：決定方式$d_6$がミニマックス決定方式になる。<br>S：平面図上で考えると、原点を１つの頂点とする正方形の右上の曲線がミニマックス決定方式になる。</td><td>・各決定方式の一般リスクの内、最大になるものを比較し、その中の最小を用いることを確認する。最小になるものを比較するのではないことに注意する。<br>・前時と同様、Geogebra を用いて、幾何学的にミニマックス決定方式を求める。</td></tr>
<tr><td>３．ベイズ決定方式を説明し、具体例で確認する。<br>S：決定方式$d_3$がベイズ決定方式になる。<br>S：平面図上で考えると、ベイズ決定方式となる直線と接するリスク点がベイズ決定方式となる。<br>S：直線の傾きは、事前確率によって求められる。</td><td>・ベイズ決定方式は、事前確率と一般リスクの期待値(平均)で求めることができることを確認する。<br>・ミニマックス決定方式と同様に、Geogebra を用いて求める。</td></tr>
<tr><td>４．事後期待損失最小化の定理を証明する</td><td>・事後期待損失最小化の定理により、ベイズ理論におけるリスクが最小であることを保証しているということを確認する。</td></tr>
<tr><td colspan="2">ミニマックス決定方式は、各決定方式の一般リスクが最大のもの同士を比較し、その中で最小となるものを選び、ベイズ決定方式は、一般リスクと事前確率の期待値が最小となるものを選ぶ。また、事後期待損失最小化の定理によって、ベイズ理論のリスクが最小であることを保証している。</td></tr>
</table>

# 引用・参考文献

(1) 赤池弘次 (1980)「エントロピーとモデルの尤度」『日本物理学会誌』,35(7):608-614
(2) 赤池弘次 (1982)「統計とエントロピー」『数学セミナー』,21(12):2-12
(3) 小島寛之 (2016)『ベイズ統計学入門』　ダイヤモンド社
(4) 繁桝算男 (1985)『ベイズ統計入門』　東京大学出版会
(5) シャロン・バーチェ・マグレイン (富永星訳)(2013)『異端のベイズ統計学』　草思社
(6) 中妻照雄 (2007)『入門ベイズ統計学』　朝倉書店
(7) 日本学術会議 (2015)『大学教育の分野別保証のための教育課程編成上の参照基準　統計学分野』
(8) 藤本熙 (1968)『統計数理の基礎と応用』　日刊工業新聞社
(9) 松原望 (2001)『意思決定の基礎』　朝倉書店
(10) 松原望 (2008)『入門ベイズ統計』　東京図書
(11) 松原望 (2010)『ベイズ統計学概説』　培風館
(12) 村上征勝 (1985)『工業統計学』　朝倉書店
(13) 文部科学省 (2008)『中学校学習指導要領解説　数学編』
(14) 文部科学省 (2009)『高等学校学習指導要領解説　数学編　理数編』
(15) 文部科学省 (2016)『平成 28 年度学校基本調査について』
(16) 文部科学省 (2016)『高等学校の教育課程に関する基礎資料』
(17) 涌井良幸, 涌井貞美 (2012)『これならわかる！ベイズ統計学』 ナツメ社