ベイズ原理はどうして生まれたのか？

植野真臣
電気通信大学

ベイズの定理

- たがいに背反な事象 $A_1, A_2, \dots, A_n$ が全事象 Ω を分割しているとする.
 - このとき、事象 $B \in \mathcal{A}$ について、
- $$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^n P(A_j)P(B|A_j)}$$
- が成り立つ.

トーマス ベイズ

• 1702-1761 ロンドン

ベイズはなぜ確率論に挑んだのか？

- 1748年に哲学者デビッド・ヒュームが人間知性研究を出版し、イエスの復活を見たという人々の主張が不正確である確率は、それがそもそも起きた確率よりも遥かに大きいと主張した。これは長老派の牧師ベイズにとって受け入れ難いものであった。ヒュームが間違っていることを証明しようと、ベイズは出来事の発生確率の定量化を試みる。

ベイズの手法

自分の背後の平らな机にボールが投げられることを考えてみよう。その着地点を推測することはできるが、実際に見ない限り、自分がどの程度正しいかを確実に知る術は無い。では、同僚がもう一つのボールを机に投げて、最初のボールの右もしくは左に着地したことを伝えた場合はどうか。例えば右に着地したのだとすれば、最初のボールは机の左側に着地した可能性が高くなる（その仮定の下では、第二のボールが右側に着地する余地が大きくなる）。同僚が新しいボールを投げるたびに、自分の推測を更新し、最初のボールの位置についての自分のモデルを改善することができる。

同様に、キリストの復活に関する様々な証言は、ヒュームが主張したような形でその出来事の信憑性を割り引くことができなことを示している、とベイズは考えた。

ベイズの結果は、たとえ弱いあいまいな証拠であっても積み重ねれば、あり得ないような出来事の低い確率を覆すことができ、それを事実として確立することができる、という手法を示した。

ベイズの手法は

- 客観的なデータから自分の信念を変える手法
- 事前確率：データのない確率、尤度：データの起こる確率
- 事後確率：データによって更新された確率
- 事後確率 $\propto$ 事前確率 $\times$ 尤度
- 世界最初のオンライン学習（逐次学習）

数学者からは不人気！！

- 主観を扱っている！！
- 事前の確率が分からない場合は、すべての可能性を等確率にする
- テーブルのどこにボールが落ちたかの最初の事前確率は一様分布。
- →
- 無知を確率で計量化しており、おかしい！！
- さらに、数学的記述が現在のものではなく、不正確なものであった！！

プライス牧師

1767年にベイズの友人だったプライス牧師は「キリスト教の重要性、その証拠、およびそれに対し申し立てられた異議」を出版し、ベイズの考えを用いてヒュームの主張に挑戦した。統計学者スティーブン・スティグラーは、プライス論文の「基本的な確率論的な論点は、ヒュームは奇跡を目撃したという独立した証言が数多くあることの重要性を過小評価したが、ベイズの結果は、たとえ危うい証拠であっても積み重ねれば、あり得ないような出来事の低い確率を覆すことができ、それを事実として確立することができる、という手法を示した」と言う。

- プライスにより、ベイズの考え方は世に出た！！
- この時点では、まだ数学的に厳密な記述はなかった。

ピエールシモン ラプラス

- 1749-1827
- フランス 数学者、物理学者

ベイズの定理の発見

- ベイズの論文から10年がたったころ、天文学の観測でなぜデータがばらつくのかの原因を推定しようとする。25歳のラプラスは以下のベイズの定理によく似た定理を導く。
- 彼自身は、原因確率の定理と呼んでいた。

$$P(A_i|B) = \frac{P(B|A_i)}{\sum_{i=1}^n P(B|A_i)}$$

ラプラス、ベイズを知る！！

- 1781年に、プライスがパリを訪れて、ベイズの話を講演した。
- その講演を聞いたラプラスは感動し、自分の作った原因確率の定理に、解釈を与え、さらに事前確率を等確率にする制約を組み込んだ。

1814年、現在のベイズの定理を導く

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i)P(B|A_i)}$$

- 地球科学、気圧変動の推定に用いる。
- ルイ18世は、ラプラスを永代伯爵の称号を譲る

ラプラス批判

- ラプラスの死後、フランス革命が起き、多くの学者たちはラプラス批判を繰り返した。ドモルガンは、ラプラスは人の論文を盗用している、と書いている。この噂はそれから150年間信じられてきた。ジョン スチュアート ミルやマリはラプラスは政治家に媚びへつらい、民衆から搾取してきたと書いている。
- 数学的理論への批判というより、フランス革命による王族への反抗であり、それに巻き込まれた形になってしまったのである。

頻度主義の台頭

- 数学者ポアソン、統計学者 カール ピアソンは、ラプラスの手法は 誤用 とでも正しい手法ではないと批判している。
- 特に 事前確率や主観確率について 強く批判している。
- ベイズの定理は絶対にはいけないもので葬り去るものであると統計学の世界では信じられてしまった。
- 特に統計学を開いたカールピアソンとロナルドフィッシャーはベイズを忌まわしいものと信じ切っており、それをを用いる人は統計学者ではないと断言していた。

頻度論の哲学

- 理論的に何度でも繰り返せる出来事だけを統計の対象とし、サンプルだけが唯一の情報源で、新たに得られたデータは、それぞれ別の問題とみるべきであると考えていた。また、統計的に判断ができるだけの十分なデータ数があればよいが、そうでないのであればそのデータはすべて廃棄すべきであると考えていた。

ロナルド フィッシャー

- 1890-1962
- イギリス、ケンブリッジ大学
- 実験計画法、分散分析、最尤法、情報量、など現在 統計学の礎を築く。
- 大のベイズ嫌いで、ベイズの普及を遅らせた張本人。ベイズの定理は誤りであると強く主張。 ベイズ学者から見ると大悪役！！

ケンブリッジのベイズ統計学者
ハロルド ジェフリーズ

- 1891-1989
- ケンブリッジ大学
- 地球物理学者、ベイズ統計学者
- 非常に穏やかな人格者
- フィッシャーと私生活では仲良いベイズ統計学者
- ジェフリーズは尤度を事前確率が一様分布のベイズの一つと考えていたため

ジェフリーズの頻度論批判

- フィッシャーが導入した有意水準、またはP値について大きな疑問があった。なぜ、実際に起きていない結果をよりどころに、仮説が棄却されて、捨て去らねばならないのだろうか？一つの実験を何度もランダムに繰り返すというが、そんなことできない場合が多い。ジェフリーズは地球物理学者であり、特定の地震を繰り返すことは無理じゃないか？頻度論学者は、データを仮説を棄却するためにのみ用いて、仮説の正しさを証明するためには用いることができないと批判。ベイズでは、データに基づいた仮説の事後確率を直接計算することができるのに。

学問上の対立

- 王立統計協会のジャーナルで二人のやり取りが何度も掲載されたが、大衆にはフィッシャーが受け入れられた。そもそもジェフリーズは温和で人を攻撃しないが、フィッシャーはその逆でわかりやすい。また、フィッシャーの手法は数式が少ないが、ジェフリーズの手法は複雑で難しい。さらにジェフリーズは話すのが不得意で、ジェフリーズが正しいときも、他の人には負けているように見えたといわれている。

ベイズ統計の没落

- 1939年 フィッシャーらの活躍で、ベイズは事実上 統計学の中で、タブーになる

軍隊で生き続けたベイズの定理

- ラプラス以降、フランス軍は、敵の位置や空気の密度や風の方向、大砲の誤差や初速などの多くの情報を入力して、照準の決め方をベイズの定理により推定した表を学ばせていた。第一次世界大戦中は、弾薬の品質管理にベイズの定理を用いていた。第二次世界大戦中にも続けられ、オペレーションズ リサーチと呼ばれるようになった。

第二次世界大戦で大活躍

- 第二次世界大戦中ドイツ軍のUボートが大きな戦果をあげ、そのための暗号 エニグマを解読することが連合国の共通の目標となる。
- 英国首相 チャーチルからの指令でエニグマを解くための数学者が集められる

ケンブリッジのあの人も

- 反ベイズの旗手ロナルド フィッシャーも手を挙げたが、ドイツの学者たちと親しい関係にあり、却下された。

選ばれたのは

選ばれたのは

- アラン チューリング
- 1912- 1954
- ケンブリッジ大学
- 数学者、コンピュータ科学者、暗号解読

- ベイズを用いて暗号の原文を確率的に推定する機械を開発する
- ばらばらに入ってくるデータを逐次的に処理するオンライン学習機

コロッセス (1943)
エニグマ解読器 (ベイズ復号機)

アメリカからの支援

- 連合軍から、アメリカの研究者が協力者に加わった。

クロード シャノン

- 1916-2001
- 米国 ベル研究所
- 情報理論の創始者
- 当時の情報理論の中核であった伝達情報量、相互情報量はベイズの事前分布の情報量と事後分布の情報量の差異のこと。すなわち、データを得ることによってどれほど情報を得ることができるかということを示している。

暗号理論

- ベイズ復号機には、情報の正しさを評価するために、事後確率と事前確率の比に対数をとったベイズ因子が用いられるようになった。
 - シャノンも 第二次世界大戦のプロジェクトのためにベイズ理論を多用している。
 - 情報理論は不確かさを減少させることが目的なのに対して、暗号理論は、復号可能な状態で不確かさをいかに増やすかという研究分野となる。
- シャノンは音声スクランブルを完成し、電話機の盗聴を困難にした。

Uボートの壊滅

- チューリングのベイズ復号機により、暗号解読に成功。
- ベイズの考えたボールの探索方式により、潜水艦の場所の特定に成功
- ドイツの潜水艦Uボートは壊滅状態に

ロシアでも軍事にベイズが大活躍

- 1941年、ヒトラーは全兵力の三分の二を使ってロシアに侵攻し、モスクワに容赦ない攻撃を開始した。ロシア軍はこの状況を打開するために カザンに疎開していたある学者に協力を要請する。

アンドレイ・コルモゴロフ

1903-1987
 ロシア モスクワ大学
 数学者 確率論、位相幾何学
 公理主義的確率論の立脚と確率論の創始者

コルモゴロフとベイズ

- ロシアの砲術では、フランス同様に 昔からベイズの射表（火砲の照準計算のための表、仰角を何度に設定すれば良いかが分かる表）が使われてきた。
- コルモゴロフは、事前分布に一様分布を用いるのがよいことを提唱した。散弾銃のように少しばらつきをもたらしたほうが、より命中率が上がると提唱した。射撃のばらつきと事前分布に関する研究が三本発表されている。

1945年 ベイズにドイツ敗戦

ドイツ軍は、連合軍でのベイズ研究により、見事 敗戦。

しかし、その後ベイズは軍事機密として、まだ学界には君臨できなくなる。

Uボート壊滅とドイツの敗戦の後

- 1952年、無名の救国の英雄チューリングは、同性愛者であることが見つかかり、英国に逮捕され、強制的に女性ホルモンを打たれた。(当時は同性愛)
- その2年後、1954年、チューリングは青酸カリをリンゴに塗り、自殺。

再評価

1. 1966年から、コンピュータ科学者らによる国際的学会のACMは、同学会の守備範囲であるコンピュータ科学を中心とした分野の最高の賞として、チューリング賞を授与している。物理や化学といったようなかなり広い分野の最高の賞、という位置づけにあるものとして、コンピュータ科学分野におけるノーベル賞に相当するものと一般に扱われている
2. 2011年12月、William Jones はイギリス政府に対してアラン・チューリングの罪を免罪（名誉回復）してほしいという電子請願を申請した。この請願には21,000以上の署名が集まったが、法務大臣はチューリングが有罪宣告されたことは遺憾だが、当時の法律に則った正当な行為であったとしてこれを拒否した。その後、2013年12月24日にエリザベス女王の名をもって正式に恩赦が発行され、キャメロン首相は、彼の業績をたたえる声明を発表した。

変人？チューリング

6月の第1週には毎年花粉症に悩まされるので、彼は花粉を吸わないようガスマスクをして自転車でオフィスに通っていた。自転車は故障していて、定期的にチェーンが外れていた。それを修理してもらおう代わりに、ペダルをこいた回数を数えて、危なくなると一旦降りてチェーンを調整していた。

マグカップが盗まれるのを防ぐために、それをラジエータパイプに鎖で繋いでいた。

第二次世界大戦に大貢献のベイズ理論

その後、どうなった？

第二次世界大戦に大貢献のベイズ理論

ドイツが降伏した数日後、チャーチルは暗号解読に関する極秘資料、文章、あらゆる証拠を消去せよと命じる。関係者には、この事実が超機密事項として伝えられた。

このことにより、ベイズ理論が第二次世界大戦の終結の大貢献した事実そのものの葬り去られることになった。

実は、この後もベイズを用いて、ソビエトの暗号を解読していたので、その手法が他国にばれることは戦略的にまずかったのである。

世界最初の汎用コンピュータ

フォン ノイマン(米) 1945年、ノイマン型コンピュータ
 モークリー、エッカート(米) 1946年、ENIAC (弾道計算)
 10進数の使用が可能で
 プログラミングが可能

チューリングは死んだ。しかし、

チューリングは死んだ。しかし

Irving John ("I. J."; "Jack") Good

1916~2009

オックスフォード大学のベイズ統計学者

900本のベイズ統計の論文を書く

世界最初の機械がオンラインで手法をベイズオンライン学習する機械学習の論文を執筆（チューリングとの仕事の延長）

レオナード・ジミー・サヴェッジ

1917-1971

シカゴ大学、コロンビア大学
主観確率の数学亭理論の構築

「アメリカが5年以内に戦争する確率」
「核兵器が使用される確率」
「ギャンブルの意思決定」

米国のベイズ主義普及のキーマン

デニス・V・リンドレー

191923- 2013

ケンブリッジ大学

現代ベイズ理論の数学的体系化を行う

ベイズ統計における重要な数学的発見

英国のベイズ派普及のキーマン

保険業界

- 統計学を学んできたアーサー L. ベイリーは、ハンコック相互保険生命の保険数理士となった。彼はドライバーの年齢、性別、事故履歴や飲酒歴などをエビデンスとして、事故確率をベイズ推定し、適正な保険料を予測するシステムを確立した。事故率は事前分布がないと正しく、推定できないことがわかった。
- これはアメリカの保険業界に革新をもたらす。保険分野ではベイズが主流になる。

1960年前半、主観確率的ベイズの提唱者、レオナード・ジミー・サヴェッジ が世界にベイリーの偉業を伝える。

物理学

エンリコ・フェルミ は ベイズにより物理モデルを発表（ノーベル賞受賞）

リチャード・ファインマン（ノーベル賞学者） スペースシャトルの事故確率をベイズによって推定

医学

- ジェローム・コーンフィールドにより、肺がんの原因をベイズで推定し、遺伝ではなく喫煙がその原因であることを証明した。また、
- 彼はコレステロールが心臓病の原因になることをベイズで示した。
- タバコ会社のコンサルタントをしていたフィッシャーは、このタバコが癌の原因であるという結果を厳しく批判。

科学と主観主義

- 「事前の意見が科学者によって違ってよいとなるとデータ解析の科学的客観性はいったいどうなるのか。」
- サヴェッジ
- 「データや証拠が集まってくると科学者たちの意見が一致してくるのと同様に、データが乏しいときには主観主義者でありつづけるが、データが多くなってくると、結果は客観的に一つに収束してくる、そして客観主義者になる。科学はまさにそうやって行われる。」

意思決定理論

- ライファとシュレイファー
- ベイズ決定木の開発
- 経済学や経営学の分野で大ブレイク

選挙の出口調査

1960年にジョン デューキーが選挙の出口調査での得票確率をベイズを用いて推論する手法を開発
事前分布に加え、性別や年齢などの補助情報も加えるので驚くべき、精度で予測できることがわかり、一般的な手法となる。

原発事故の確率

- ラムッセンが1974年に原発事故の起こる確率をベイズ推定。高い確率で事故が起ると予測していた。
- 頻度派が大きく、この結果に反対する。
- 1979年に スリーマイル島の原発事故が起こる。

1980年以降コンピュータの普及

- 頻度派の統計学は 紙と鉛筆で計算できるように考えられてきた。
- ベイズは複雑な数理であるが、コンピュータの普及に伴い、実現化してくる。

スペースシャトル チャレンジャーの事故確率

- 1983年にアメリカ空軍が契約した統計分析会社 テレデザイン社がベイズにより、スペースシャトル チャレンジャーの事故の確率を推定したところ、頻度派のコンサルタントが出した結果 10万分の1とは異なり、1/35と推定。
- 頻度派はベイズの結果に大きく反対する。
- 1986年チャレンジャーは25回目の打ち上げで大爆発を起こす。

ディーゼルエンジンが肺がんを与える影響分析

- 1983年デュムシェルは、ディーゼルエンジンの肺がんへの影響はむしろできることを証明。

ベイズ画像解析

- 1985年
- スチュアート ジーマンとドナルド ジーマン が画像解析のためにMCMC法（マルコフチェーンモンテカルロ法）によるギブスサンプリングアルゴリズムを開発。

経済学

- ゲーム理論のベイズ学者 ナッシュがナッシュ均衡で1994年ノーベル賞
- ハリーマーコウィッツ 2002年にベイズ理論経済学でノーベル賞
- カーネマンとトベルスキー 実際の人間の意思決定理論ではベイズを少し崩さないといけないことを発見
- 2002年にノーベル賞

1990年

- デヴィッド ヘッカーマン
- スタンフォード大学 医学部終了後、マイクロソフト社へ
- ベイジアンネットワークの教祖
- ジュディア パールは 命名したこととベイジアンネットワークの前の確率木の開発者

MS商品

- 手書き文字認証
- お勧めシステム
- HELPの対応システム
- スпамメール駆除
- 売り上げ予想システム
- 交通渋滞の予想システム
- ペーパークリップのキャラクター クリッピー

グーグル

- 検索エンジンにベイズ
- スпамメールなどの処理に ナイブベイズ

推薦システム

- ネットフリックスが主催した10億円のコンテストで2009年にA T & T社のベイジアンネットワークによる推薦システムが優勝

自然言語処理

- マーサーとブラウン
- 1998年 IBMベイズによる機械翻訳システム
- 2005年 グーグル ベイズ機械翻訳システムが世界機械翻訳コンテストで優勝
- 現在、グーグルは25言語でベイズ翻訳システムを実用化

アダプティブ ラーニング

ACT*の開発者であるカーネギーメロン大学の人工知能学者アンダーソンのグループは、ベイジアンネットワークで学習者の行き詰まりを同定し、適正な問題を選択し、適正なヒントを提示するシステムを1990年代に提案している。その後、商品化され、現在ではKnewton社が多くのシステムを開発し販売している。2016年より日本にも進出してくる。

<http://wired.jp/2013/04/27/adaptive-learning/>

- **一人ひとりにあった学習を実現！ 教育業界の新潮流「アダプティブラーニング」**
- 個人個人に最適化された学習内容の自動提供を実現する「アダプティブラーニング」。ビッグデータを用いたその仕組みによって、これから教育分野でGoogleに匹敵する企業が誕生する可能性もあると期待されている。その注目の動向を紹介しよう。

まとめ

1. ベイズの考え方は、キリスト教の牧師ベイズがキリストの復活の目撃者の証拠により復活があったことを証明しようとしたことが始まり、数学的な提案者は数学者ラプラス。
2. 事前分布、主観確率のために統計学の分野では長く毛嫌いされてきた。
3. 第二次世界大戦を終わらせたのはチューリングのベイズ機械学習。
4. 戦後、グッド、サヴェッジ、リンドレーの3人がベイズを普及
5. 1980年代にジーマン兄弟のIEEE論文、MCMCにより一機に普及
6. 1990年以降ベイジアンネットワークが出現し、マイクロソフト、グーグル、IBMなどで商品開発され、様々な分野で活躍