

ベイズはコンピュータ、 人工知能の父である!!

植野真臣
電気通信大学
情報理工学研究科
情報数理工学プログラム

今後のスケジュール (予定)

4月10日	授業の概要とガイダンス
4月17日	ベイズの定理
4月24日	ベイズはどのように誕生したか?
5月1日	ベイズはコンピュータ、人工知能の父である!!
5月8日	アランチューリングとベイズ
5月15日	ベイズから機械学習へ
5月22日	確率の基礎の復習
5月29日	ベリフとベイズ
6月12日	尤度と最尤推定
6月19日	数値計算法による推定
6月26日	ベイズ推定と事前分布
7月3日	マルコフチェーンモンテカルロ(MCMC)法
7月10日	ベイジアンネットワーク
7月17日	ベイジアンネットワークと機械学習
7月24日	テストと総括

授業の目標

ベイズの定理とコンピュータ誕生の
関係を学ぶ!!

1814年、現在の定理を導く

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

地球科学、気圧変動の推定に用いる。
ラプラスはナポレオンにかわいがられ、
内務大臣も務めた。

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 とて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用い
て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。

簡単に検定を説明しましょう！！

検定

薬Dは血糖値を下げるのに有効かどうかをどうやって確かめるか？

検定

薬Dは血糖値を下げるのに有効かどうかをどうやって確かめるか？
患者からランダムに二つのグループA,Bを作り、Aには薬Dを投与し、Bには何もしない。実験の最初と最後に血糖値を測定し、各グループの患者の血糖値の差の平均を求めて2グループで比較すればよい。Aの平均値のほうがBより優位に大きければ良い。

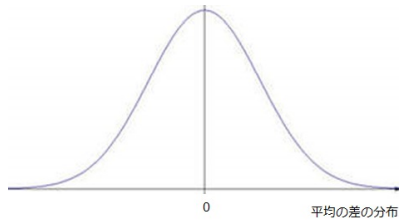
仮説

- 帰無仮説：棄却されることを期待して立てる仮説「2グループの平均に差がない」
- 対立仮説：示したい仮説「2グループの平均に差がある」

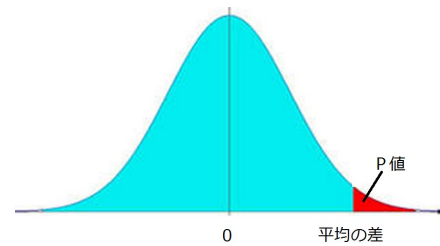
統計的検定では

帰無仮説を元に理論を作ります！！

2グループが同一正規母集団から発生しているときの差異の分布



誤り確率をP値



検定とは

有意水準 $\alpha = 0.05$ などを設定し、Pが α 以下のときに仮説が棄却され、有意差があったと判定できる。

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。帰無仮説に従ってモデルをつくり、データがそのモデルに当てはまる確率がP値である。棄却したいのでP値が小さいと棄却できる。

ジェフリーズの頻度論批判

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、そんなことはできない。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。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。

ベイズ V.S. 統計検定

- H 仮説、 $\neg H$ 帰無仮説 (棄却したい仮説)
- X データ
- ベイズ的考え方

$$P(H|X) = \frac{P(X|H)P(H)}{P(X|H)P(H) + P(X|\neg H)P(\neg H)}$$

- 統計的検定の考え方

$P(X|\neg H)$ がある値より小さいと $P(H|X)$ が大きいと考える (根拠薄い)

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。

BOMBÉ (世界で最初のコンピュータ)

- 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

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。

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ファクター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。

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

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。
2. 2013年12月24日にエリザベル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。

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う代わりに、ペダルをこいだ回数を数えて、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。
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。

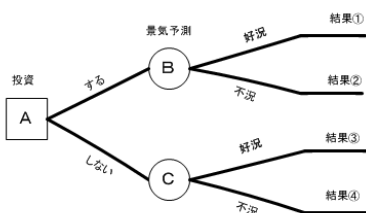
保険業界

- 統計学を学んできたアーサー L. ベイリーは、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。

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。」

デシジョンツリー

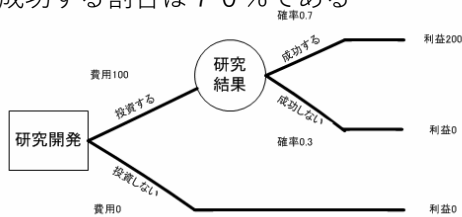


- 意思決定ノード：意思決定者がコントロールできる行動で、「□」で表わす。
- イベントノード：意思決定者がコントロールできない事象で「○」で表わす。
- 結果ノード＝リンク先で表す
- 結果ノード：結果価値を得る最終点開いたリンクで表わす。

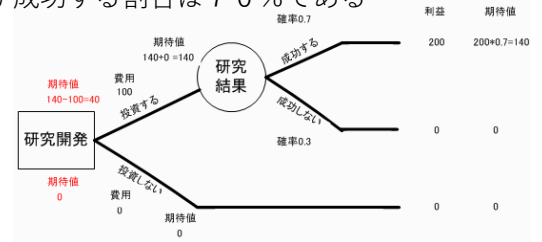
例題7 デシジョンツリーでは投資すべきか？

- 研究投資には100億円が必要
- 研究が成功すると利益が200億円見込める
- 研究が成功する割合は70%である

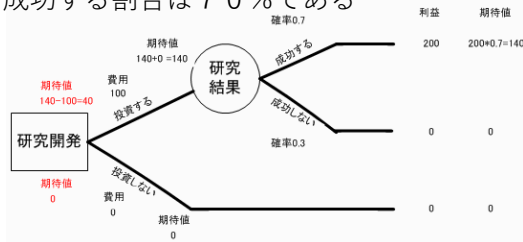
- 研究投資には100億円が必要
- 研究が成功すると利益が200億円見込める
- 研究が成功する割合は70%である



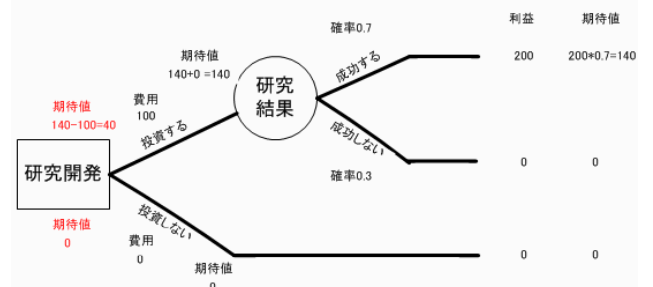
- 研究投資には100億円が必要
- 研究が成功すると利益が200億円見込める
- 研究が成功する割合は70%である



- 研究投資には100億円が必要
- 研究が成功すると利益が200億円見込める
- 研究が成功する割合は70%である



結論:投資したほうが40億円期待値が高いので投資せよ。



原発事故の確率

- ハロルド・ルイスが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こると予測していた。
- 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。

ベイズ的手法の波及

『日本の原発事故率が7倍 500年に1回(54基なら10年に1回)に改定』
 原発保険料7倍に…1か所、年2億4000万円。
 政府は20日の閣議で、原子力発電所の事故による損害賠償に備え、電力会社に義務づけている保険契約について、4月1日から電力会社が負担する保険料を約7倍に引き上げることを決めた。

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す

ディーゼルエンジンが肺がん に与える影響分析

- 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無視できることを証明。

経済学

- ゲーム理論のベイズ学者 ナッシュがベイズ・ナッシュ均衡で1994年ノーベル賞 ベイジアンゲームと呼ばれる。
- ハリーマーコウイツツ 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
- スパムメールなどの処理に ナイーブベイズ

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AT&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
- 機械学習 トップカンファレンス UAI が開催した予測コンテストで2010年 UCLA グループのベイジアンネットワークが優勝

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「アダプティヴラーニング」
- 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「アダプティヴラーニング」。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. ベイズは、牧師がキリストの復活を証明しようとしたことに始まり、数学的にラプラスが提案した。
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. 世界最初のコンピュータはチューリングのベイズ機械学習。
4. オペレーションリサーチ、情報理論、統計物理、意思決定理論、ゲーム理論などの分野につながる
5. 1980年代にジーマン兄弟のIEEE論文、MCMCにより一機にベイズが普及
6. 1990年以降ベイジアンネットワークが出現