

ベイズ統計

古谷知之

(c) Tomoyuki Furutani, Keio University SFC

1

講義概要

- 授業スケジュールと成績判定方法
- ベイズ統計を巡る歴史的背景
- 確率的（ベイズ的）意思決定
- ベイズ統計の社会的応用例

(c) Tomoyuki Furutani, Keio University SFC

2

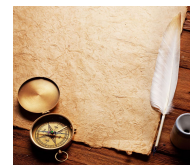
この授業の目的

- 「ベイズの定理」に基づくベイズ的思考方法を身につける
- ベイズ統計学への理解を深める
- 統計モデリングのベイズ推定に関する理論と技法を習得する

(c) Tomoyuki Furutani, Keio University SFC

3

授業スケジュール



第01回	イントロダクション	第08回	ベイズ推定 (1)
第02回	確率 (1)	第09回	ベイズ推定 (2)
第03回	確率 (2)	第10回	MCMC (1)
第04回	ベイズの定理 (1)	第11回	MCMC (2)
第05回	ベイズの定理 (2)	第12回	ベイズ統計データ分析 (1)
第06回	ベイズの定理の応用 (1)	第13回	ベイズ統計データ分析 (2)
第07回	ベイズの定理の応用 (2)	第14回	ベイズ統計データ分析 (3)

履修者の状況により授業の進度を変更することもあります。

(c) Tomoyuki Furutani, Keio University SFC

4

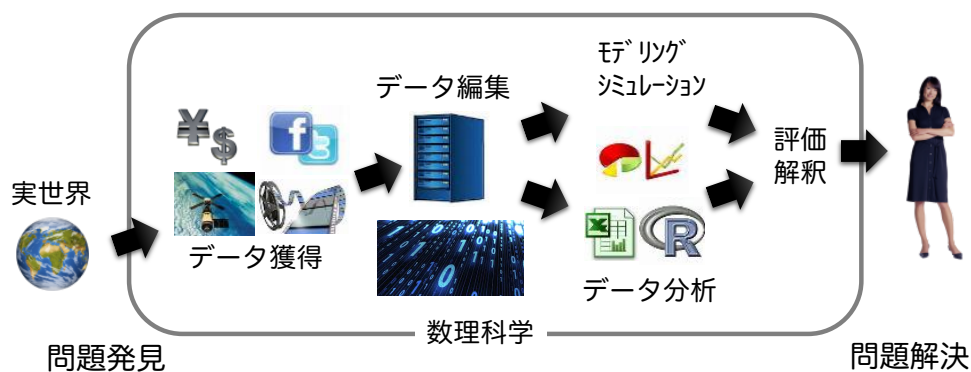
成績判定方法

- ミニレポート（各回の内容に関するクイズやR演習課題など）
- 期末レポート

(c) Tomoyuki Furutani, Keio University SFC

5

データサイエンスを活用した 問題発見・問題解決



(c) Tomoyuki Furutani, Keio University SFC

6

データ科学に期待される役割

- 政策の質向上
 - 予測・分析・モニタリング
精度改善
 - 課題解決の迅速化
- 政策サービス改善/効率化
 - 生活者・地域（患者）視点からの政策
 - パーソナル化、オーダーメイド化
 - 情報公開、プライバシー保護

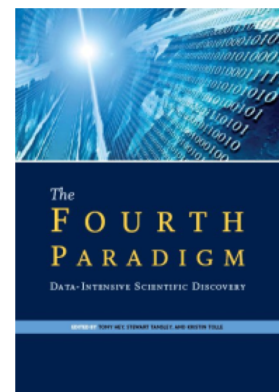


(c) Tomoyuki Furutani, Keio University SFC

7

21世紀はデータサイエンスの時代

- 「ビッグデータ」や「オープンデータ」という言葉を一度は聞いたことがあると思います
- 今まで以上に多様で膨大なデータの解析が可能になり、データサイエンスは21世紀学問の主流を占めるといわれています
- データサイエンスに精通した人材育成が世界的な課題となっています



T. Hey, S. Tansley and K. Tolle (2009) "The Fourth Paradigm: Data-Intensive Scientific Discovery", Microsoft Research.

(c) Tomoyuki Furutani, Keio University SFC

8

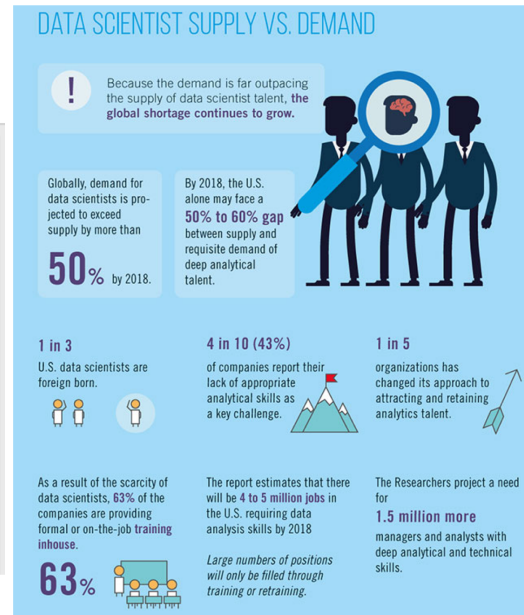
データサイエンティスト養成への強い社会的ニーズ

50 Best Jobs in America for 2019

Best Jobs | 2019 | United States

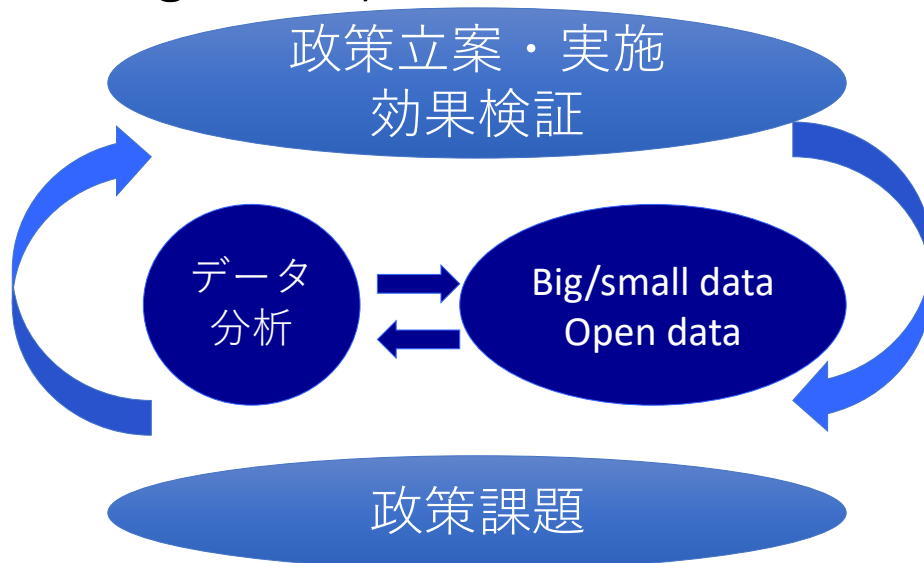
Job Title	Median Base Salary	Job Satisfaction	Job Openings
#1 Data Scientist	\$108,000	4.3/5	6,510
#2 Nursing Manager	\$83,000	4/5	13,931
#3 Marketing Manager	\$82,000	4.2/5	7,395
#4 Occupational Therapist	\$74,000	4/5	17,701
#5 Product Manager	\$115,000	3.8/5	11,884
#6 Devops Engineer	\$106,000	4.1/5	4,657
#7 Program Manager	\$87,000	3.9/5	14,753

https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm



(c) Tomoyuki Furutani, Keio University SFC <https://insidebigdata.com/2018/08/19/infographic-data-scientist-shortage/>

データ科学時代のEBPM (Evidence Based Policy Management)



(c) Tomoyuki Furutani, Keio University SFC

Thomas Bayes (1701?-1761)

- 英国非国教会派の牧師
- 死後に、初期の条件付き確率に関する論文が発見され、遺作として発表される
 - 「事後」で条件づけられた「事前」の確率を計算できること = 「ベイズの定理」を発見



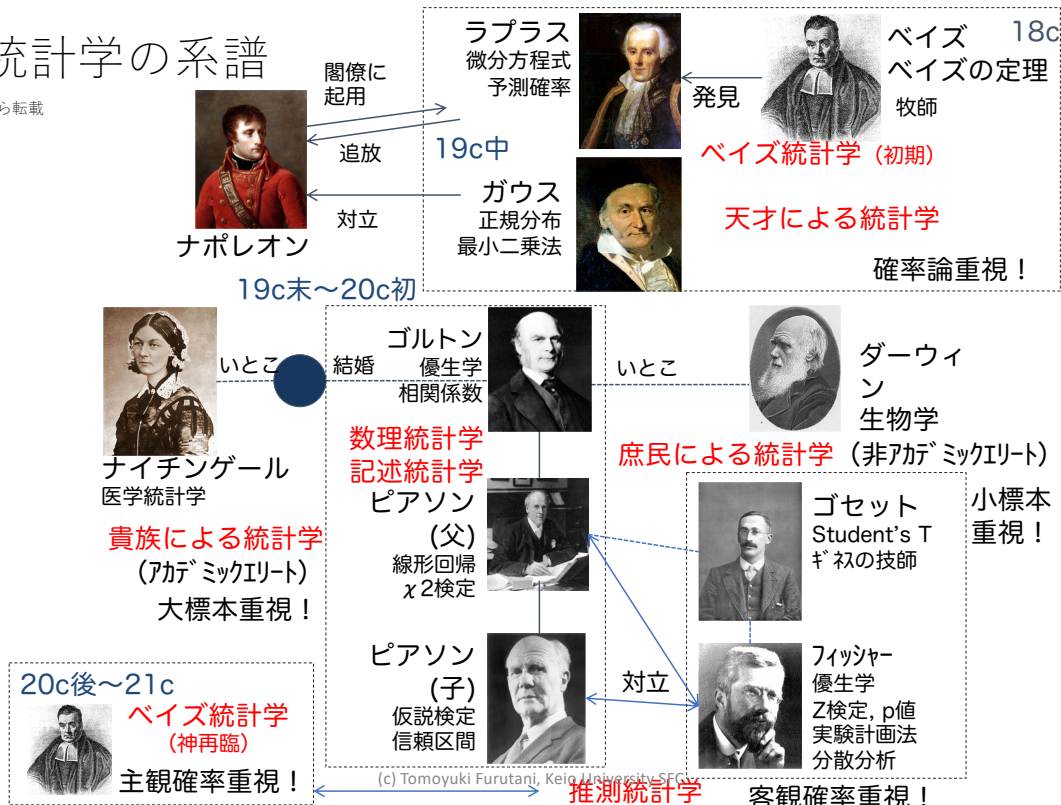
<http://plus.maths.org/>



(c) Tomoyuki Furutani, Keio University SFC

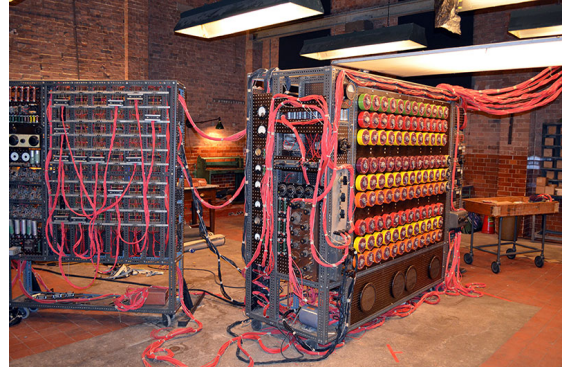
近現代統計学の系譜

画像は全てWikipediaから転載



チューリング・マシンとU-ボート

- アラン・チューリングがドイツ軍の暗号解読のための推論にベイズ理論を採用
- U-ボート撃破に成功
- 映画『イミテーション・ゲーム』
- 統計学を取り入れた軍事作戦支援ツールは、第2次世界対戦後にはOR（オペレーション・リサーチ）として発展



<http://www.dbprops.co.uk/portfolio-items/imitation-game/>

(c) Tomoyuki Furutani, Keio University SFC

15

JFKとマクナマラ

- J・F・ケネディ
 - 第35代米合衆国大統領
- ロバート・マクナマラ
 - 米国防長官。国防総省で分析革命を先導
- アンドリュー・マーシャル
 - ランド研究所の統計学者。国防戦略研究にORを活用。



画像は全てWikipediaから転載

(c) Tomoyuki Furutani, Keio University SFC

17

ベイズ統計が用いられる例

- 選挙の予測（政治学）
- 金融オペレーションリスク管理（経済学）
- 新薬の治験効果検証（医療・バイオ）
- 生物多様性の分析（生態学）
- 犯罪捜査
- スポーツ選手の行動履歴
- :
- ベイズ統計を使わないと論文が認められない分野も→ベイズさえ身につければanalystとしての職に困らない

(c) Tomoyuki Furutani, Keio University SFC

18

事前確率・事後確率・尤度

- ベイズの定理にもとづいて、 $P(H)$ を事前確率、 $P(D|H)$ を尤度、 $P(H|D)$ を事後確率という

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

事後確率 $P(H|D)$ 、尤度 $P(D|H)$ 、事前確率 $P(H)$ はそれぞれ赤い楕円で囲われ、青い矢印でラベルされている。

(c) Tomoyuki Furutani, Keio University SFC

19

テロリストを突き止めるには

- 尤度比を用いて、テロリストと一般人を含むネットワーク（グラフ）から、テロリスト含む（一般人を含まない）ネットワークの尤もらしさを抽出

$$\text{尤度比 } P_H(G) = \frac{P(G | \begin{array}{c} \text{一般人} \\ \text{テロリスト} \end{array} +)}{P(G |)}$$

(c) Tomoyuki Furutani, Keio University SFC

20

Bayesian analysis in military operations

- 洋上で紛失した武器の捜索にベイズ理論が用いられた
- 紛失した確率をベイズ理論により推測
- 近年では無人機やロボットを活用したオペレーションにベイズ統計が応用



U.S. Army Research Laboratory
<http://www.arl.army.mil/www/default.cfm?page=185>

(c) Tomoyuki Furutani, Keio University SFC

21

原子力潜水艦スコープオン号

- ベイズ推定により沈没したスコープオン号を発見

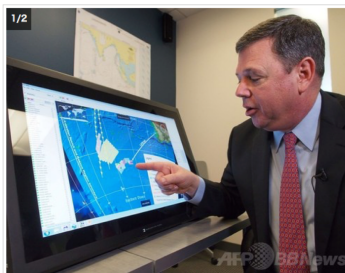


(c) Tomoyuki Furutani, Keio University SFC <https://www.youtube.com/watch?v=WH-HWWkCMeU> 22

墜落したマレーシア機の捜索にもベイズ統計が活用された

不明機捜索に「数学の魔術師」集団が名乗り

2014年03月27日 16:38 発信地：レストン/米国



米バージニア (Virginia) 州レストン (Reston) にある科学コンサルティング企業メトロン (Metron) の本社で取材に応じる同社のバン・ガーリー (Van Gurley) 氏 (2014年3月26日撮影)。
©AFP/Robert MacPherson



▶ メディア・報道関係・法人の方向け写真購入のお問合せ

【3月27日 AFP】2009年に大西洋に墜落したエールフランス (Air France) 機の残骸の位置を特定した「数学の魔術師」たちが、このほど消息を絶ったマレーシア航空 (Malaysia Airlines) MH370便の捜索にも協力する意気込みを見せている。

米国の科学コンサルティング企業メトロン (Metron) は今のところ、同機の捜索への協力要請は受けていない。だが既に、これまでに公表されているわずかなデータを基に、調査に着手した。

米首都ワシントン (Washington D.C.) に近いバージニア (Virginia) 州レストン (Reston) にあるメトロン本社で26日、AFPの取材に応じた同社の高等数学応用部門を率いるバン・ガーリー (Van Gurley) 氏は、「独立した評価を始めるため、公表されている全データの入手を試みている」と語った。評価の結果は、「関心を持つ人なら誰にでも公開する」意向だという。

1982年に設立されたメトロンには現在、応用数学の専門家らを含む170人の職員がおり、ソナー (水中音波探知機) システムなど、米国の国家安全保障に高度に特化した数理解析を行っている。

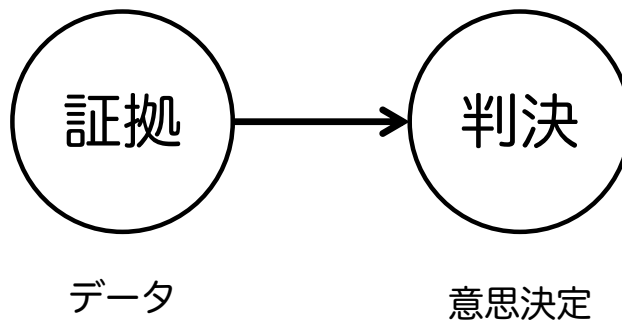
だが同時に、18世紀初期の英国の統計学者で哲学者、長老派の牧師でもあったトーマス・ベイズ (Thomas Bayes) が考案した定理に基づき同社が開発した捜索救助手帳は、米沿岸警備隊の間で幅広く利用されている。

この手帳は、「ある問題に関する情報のうち、入手可能なもの全てについて調べ、各情報に確信度 (どれだけ確信がもてるか) をつけることを強いる、構造化された手法だ」(ガーリー氏) という。この手法の下ではどんなにささいなデータも捨てられることはないが、情報が「確実になっていくにつれて (例えば、衛星画像に写った点が本物の残骸だということが判明した時など)、対象物が特定の位置にある確率が高まることになる。

<https://www.afpbb.com/articles/-/3011064>

(c)

裁判における意思決定



証拠があるという条件下で有罪にできる可能性は？

(c) Tomoyuki Furutani, Keio University SFC

27

裁判とベイズ統計

- 裁判で確率と統計的証拠が用いられる事例は増えている
- 「統計的有意性」は法的な意思決定においても重要性を増している
- ベイズ統計を用いた判例も少なくない

裁判官が法廷でベイズ統計の使用を制限～テリー・プラチェット効果に判事が立腹 (The Register)

裁判官が法廷でベイズ統計の使用を制限 ?テリー・プラチェット効果に判事が立腹

裁判官が法廷でベイズ統計の使用を制限
～テリー・プラチェット効果に判事が立腹

f シェア ツイート

もっと「The Register」のニュース

ニュージーランド乱射事件犯人、同程度の時間かけ...

クレプス氏も標的か、米男性と英10代受刑者が連...

2019年新春の決意：お年玉で絶対に買ってはいけ...

編集部にメッセージを送る

By Tim Worstall

Posted in Law, 5th October 2011 09:00 GMT

分析 英国の（悲しいことに名称のない）ある事件の裁判官が、ベイズの定理（犯人特定の確率を予測するため、裁判所で使用される手法）は、刑事裁判では使用されるべきではないと判決を下した。

少なくとも、近年見られるような依存をするべきではないという。同裁判官によれば、被告が有罪である可能性について陪審に説明するため、鑑定人がデータを同定理に当てはめるなら、それ以前に、基本となっている統計は概算推定ではなく、「確固たる」ものであるべきだからだ。この考えにより、DNAは別として、薬物の痕跡、衣類の繊維、足跡などが容疑者と適合する確率などにも影響が出る可能性がある。

<https://scan.netsecurity.ne.jp/article/2011/10/12/27436.html>

(c) Tomoyuki Furutani, Keio University SFC

28

「統計的有意性」 偏重への懐疑論

- 800名超の科学者が「統計的有意性」 偏重主義に反対署名
- 「95%水準で統計的に有意である」結果のみが重要な結果なのか？
- 「統計的に有意である/有意でない」ということが正しく理解されているか？

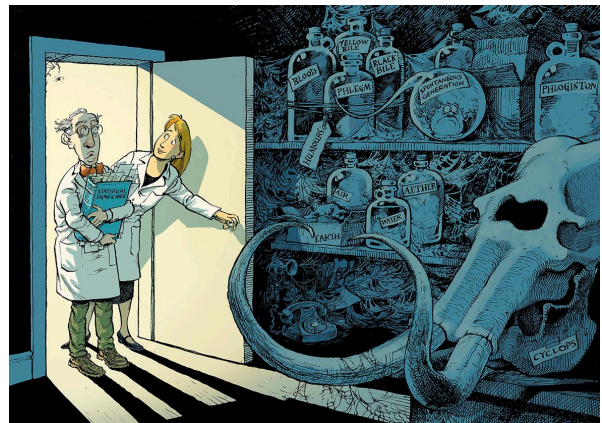


COMMENT · 20 MARCH 2019

Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein, Sander Greenland & Blake McShane



(c) Tomoyuki Furutani, Keio University SFC

<https://www.nature.com/articles/d41586-019-00857-9>

29

「統計的有意性」 偏重への懐疑論

- P値論争
 - 統計的有意性にP値を用いることが妥当なのかどうかについて、様々な批判や議論が展開されてきた
 - アメリカ統計学会（ASA）がP値の限界などについて声明発表
 - インパクトファクターが高い学術誌などがP値の使用を禁止
- ビッグデータの普及
 - 小規模標本データを前提とした統計解析手法に縛られる必要はない

(c) Tomoyuki Furutani, Keio University SFC

30

アメリカ統計学会(ASA)による声明(2016)

1. p値はデータと特定の統計モデルが矛盾する程度を示す指標の一つである
2. p値は調べている仮説が正しい確率や、データが偶然のみで得られた確率を測るものではない
3. 科学的な結論や、ビジネス、政策における決定は、p値がある値を超えたかどうかのみに基づくべきではない
4. 適正な推測のためには、全てを報告する透明性が必要である
5. p値や統計的有意性は、効果の大きさや結果の重要性を意味しない
6. p値は、それだけでは統計モデルや仮説に関するエビデンスの良い指標とはならない

日本生物学会HPより (<http://biometrics.gr.jp/news/all/ASA.pdf>)
(c) Tomoyuki Furutani, Keio University SFC

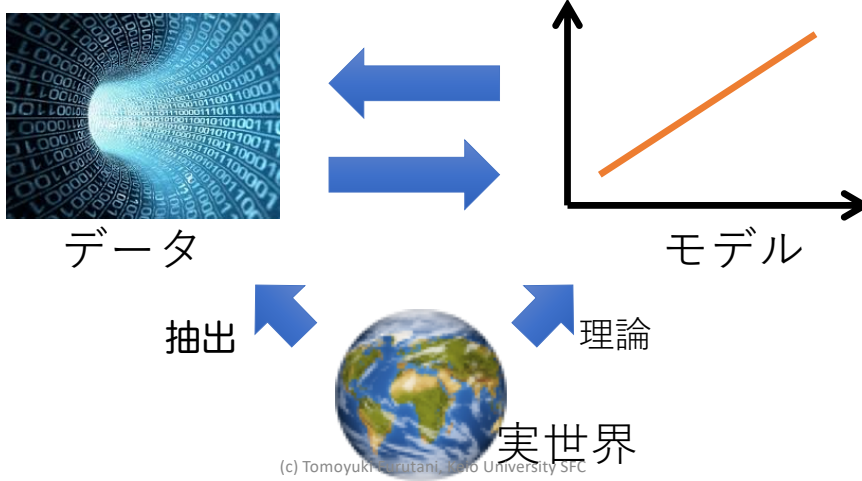
31

21世紀は確率論とベイズの時代

- 20世紀まで
 - 「データが十分に得られない時代」
 - データが少ないので数学的理論を精緻化
 - 頻度主義統計、微積分
- 21世紀
 - 「大量のデータが得られる時代」へ
 - データの分布を考えた分析手法
 - 確率論、ベイズ統計

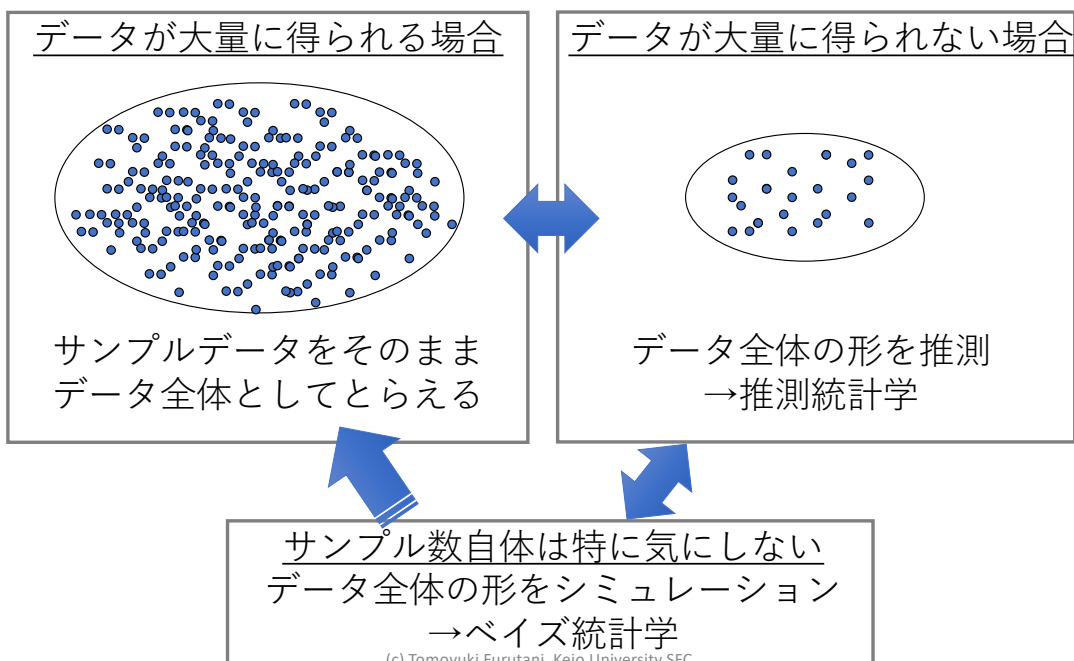
古典統計とベイズ統計の違い

古典統計（推測統計学）：データをモデルにあわせる
ベイズ統計：データにあわせて自由にモデリング



33

データサイエンスの基本的な考え方

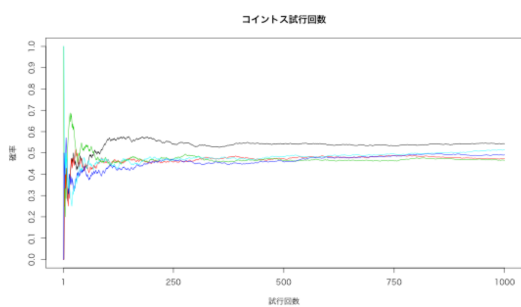


34

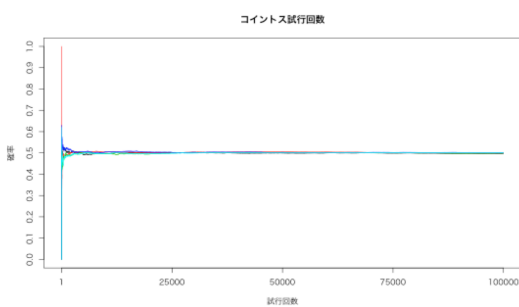
大量データを使えるようになると...

- コイントスをするとうる確率は $1/2$?

5人が1000回コイントス



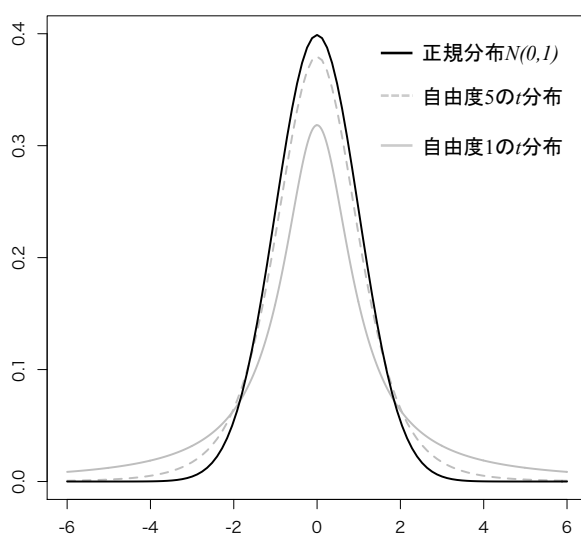
5人が100000回コイントス



(c) Tomoyuki Furutani, Keio University SFC

35

大量データを使えるようになると...



(c) Tomoyuki Furutani, Keio University SFC

36

なぜ今、ベイズ統計なのか？

- 政策効果(A)を示す政策知見とデータ(B)
- 異なる分野のデータと政策知見を、比較的大量に組み合わせることが出来る
- 不確実性を数値化（確率表現）できる

政策評価に必要なデータ収集 政策評価指標

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

政策効果の不確実性 これまでの政策とデータの蓄積 複数の分野にまたがる政策課題

(c) Tomoyuki Furutani, Keio University SFC

37

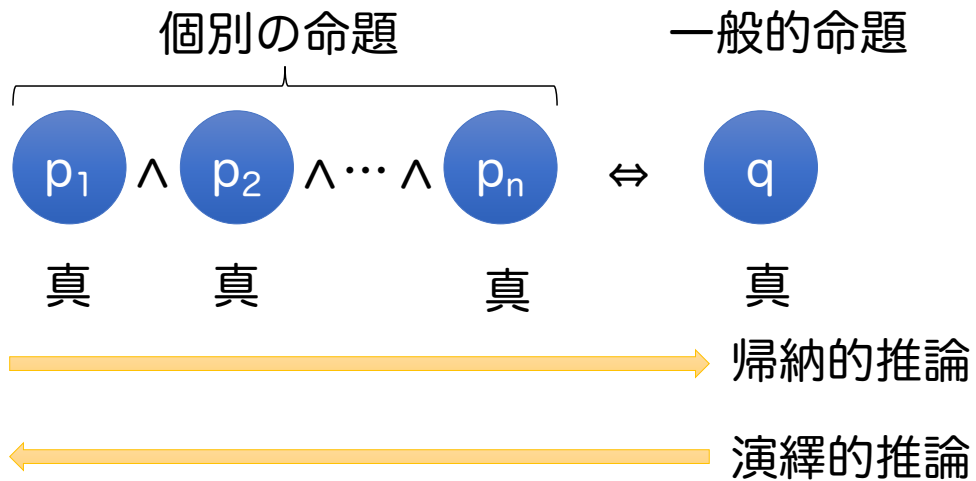
仮説（一般命題）と証拠（仮説命題）

- 仮説（一般命題）
 - ゲームに勝利する
- 証拠（個別命題）
 - チーム（自軍）は強い
 - 強い証拠A、B、...

(c) Tomoyuki Furutani, Keio University SFC

38

演繹的推論と帰納的推論



(c) Tomoyuki Furutani, Keio University SFC

39

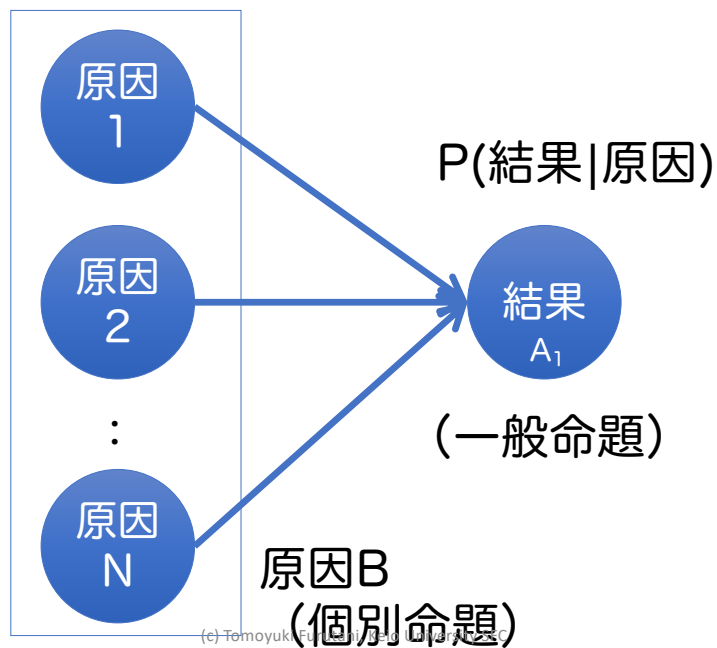
演繹的推論と帰納的推論

- 演繹的推論
 - 前提 1 : Aは犯人に違いない
 - 前提 2 : Aが犯人なら証拠がある
 - 結論 : Aに証拠が見つかる
- 帰納的推論
 - 前提 1 : Aに証拠が見つかる
 - 前提 2 : 証拠があれば犯人である
 - 結論 : Aは犯人である

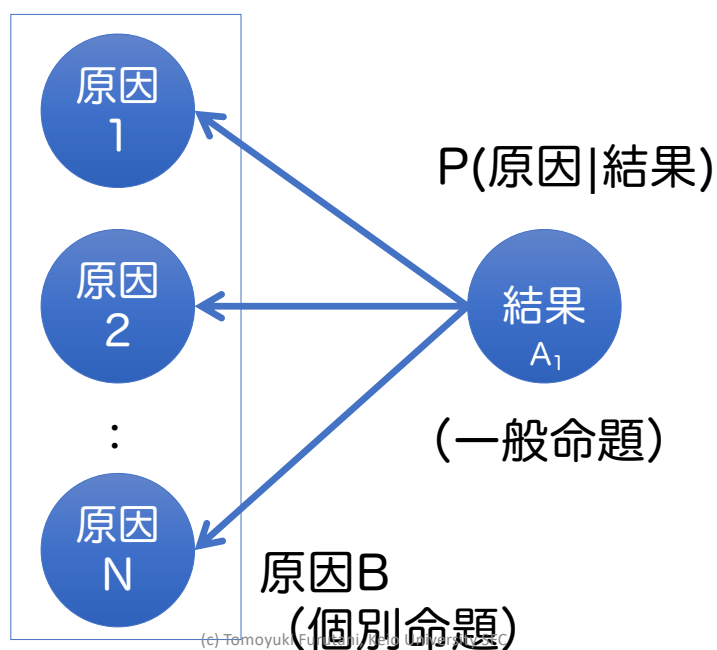
(c) Tomoyuki Furutani, Keio University SFC

40

因果関係と条件付き確率：帰納的推論



逆確率による原因の推定：演繹的推論



演繹的推論と帰納的推論

- 演繹的推論

- 「ゲームに勝利する」という仮説（命題）が正しいことを前提に、「強いチームである」という証拠（命題）が正しいことを、演繹的に証明する。

- 帰納的推論

- 「強いチームである」という証拠が正しいことを前提に、「ゲームに勝利する」という仮説が正しいことを帰納的に推論する。

戦略・戦術・技術

- 戦略

- 資源の動員と配分、配置

- 戦術

- ゲームを遂行するために発揮する能力

- 技術

- 攻撃・防御のための具体的手段

仮説・証拠・推論

- 仮説
 - ある戦略・戦術・技術を選択するとゲームに勝利できる
- 証拠
 - あるゲームに勝利した
- 推論
 - ある戦略・戦術・技術を選択するとゲームに勝利できるはずだ

帰納法的発想と演繹的发想

- 帰納法的発想
 - データを収集し仮説（モデル）の正しさを証明しようとする方法
 - 統計学
- 演繹的发想
 - 仮説が正しいことを前提に、証拠の尤もらしさを示そうとする方法
 - 確率論

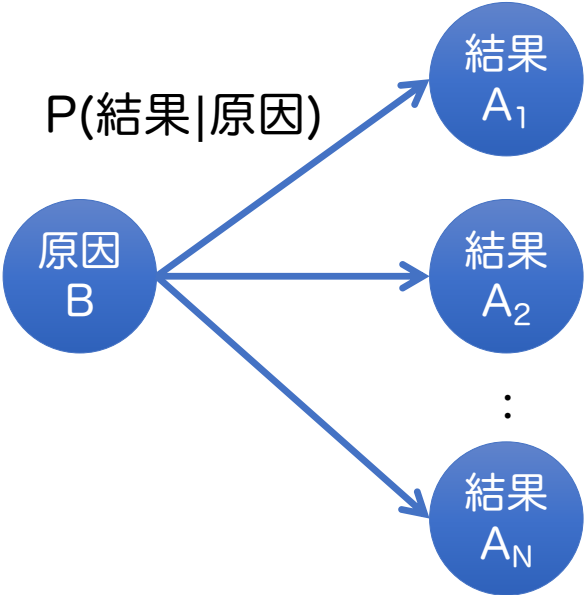
参謀の役割

- 「ゲームに勝利する」という目標のもと、「チーム力を強化する」という目的を掲げる。
- 相手チームより多くのポイントを獲得できればゲームに勝利できるとき、
 - 目的：ゲームに勝利すること
 - 命題：相手チームより多く得点すること

帰納法的発想の限界

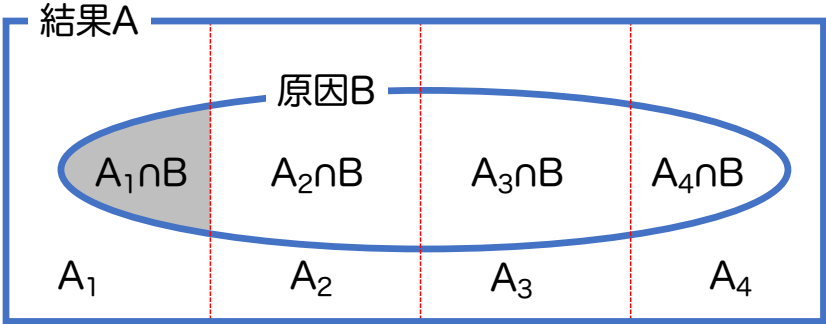
- 推論の前提となる「データ」以上の発想は得られない
- 太平洋戦争の戦略策定の方法
 - 日本軍：帰納法的手法
 - 米国軍：演繹的手法 …と言われている
- 「ゲームに勝利する」という命題に勝利するには限界がある？

原因から結果の推定



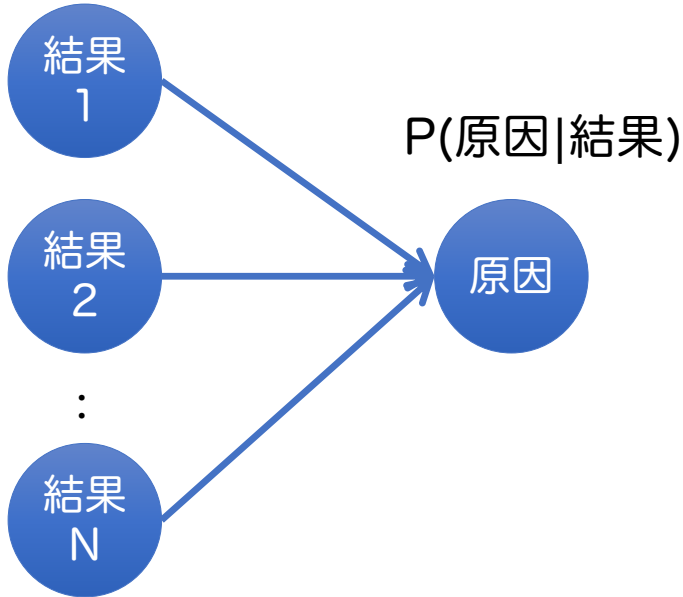
(c) Tomoyuki Furutani, Keio University SFC

原因と結果の関係



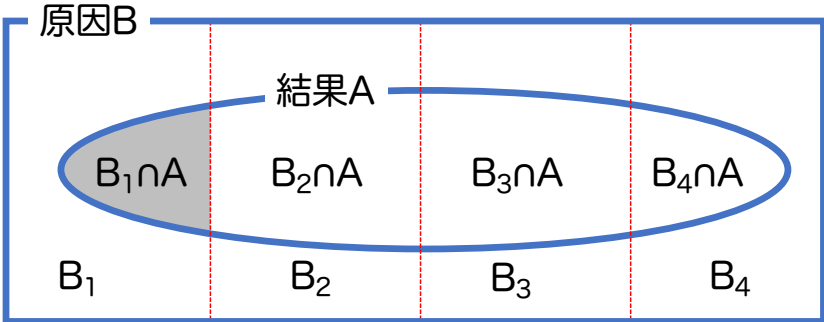
(c) Tomoyuki Furutani, Keio University SFC

逆確率による原因の推定



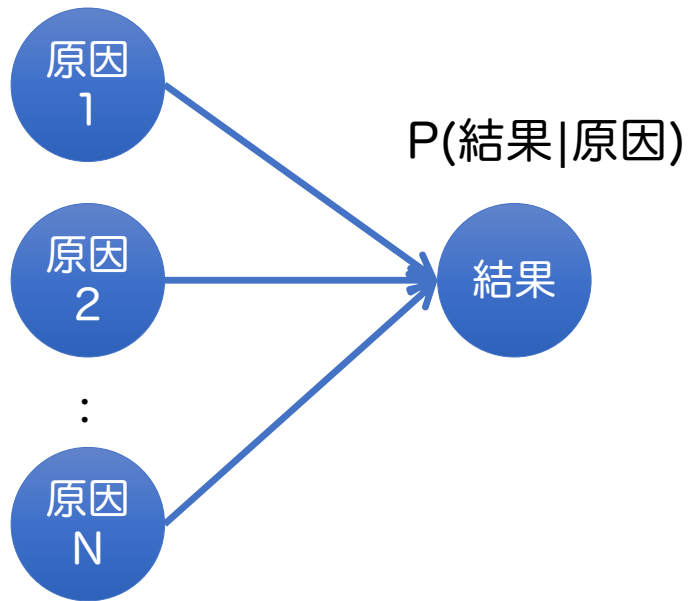
(c) Tomoyuki Furutani, Keio University SFC

逆確率における原因と結果の関係



(c) Tomoyuki Furutani, Keio University SFC

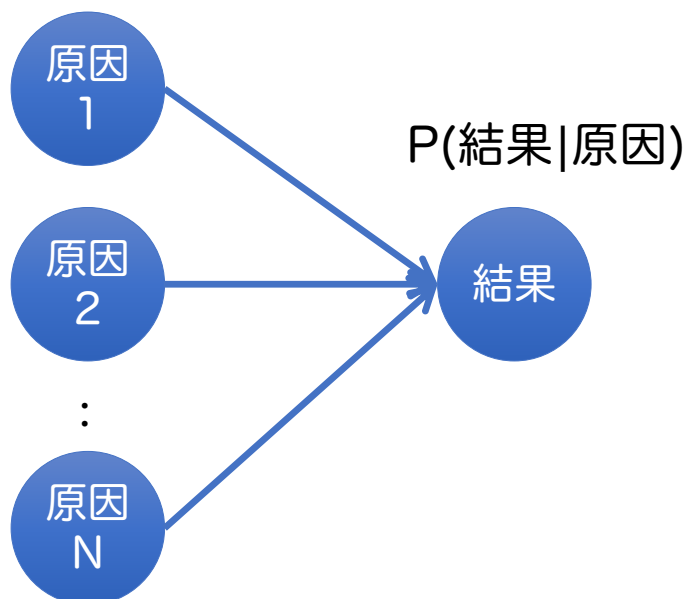
因果関係



(c) Tomoyuki Furutani, Keio University SFC

53

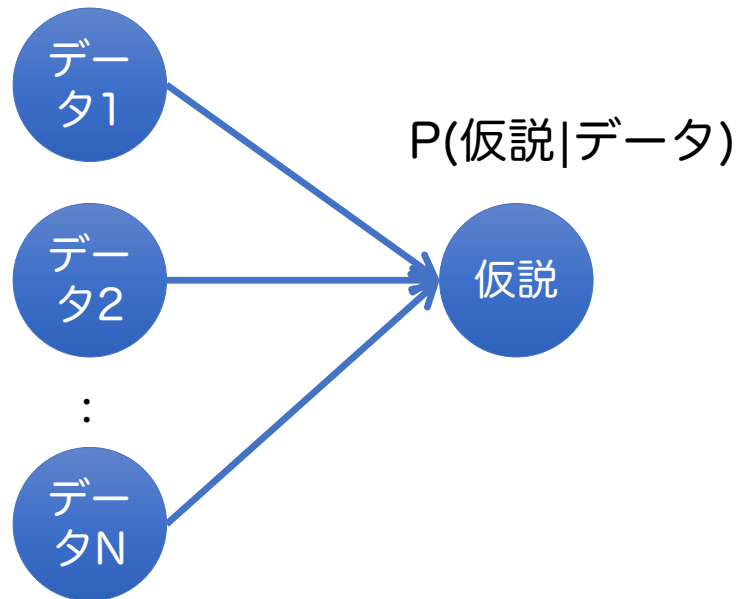
因果関係と条件付き確率



(c) Tomoyuki Furutani, Keio University SFC

54

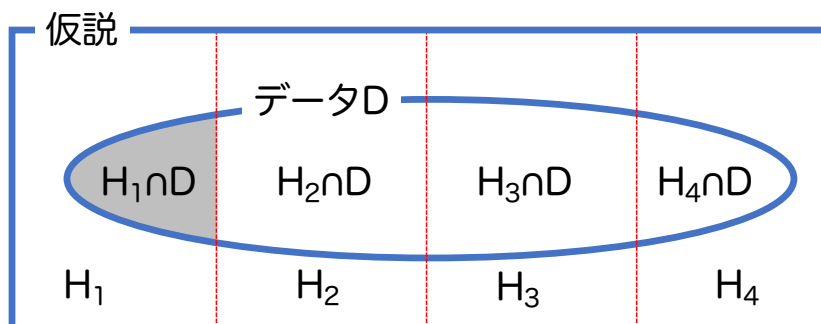
データから仮説（モデル）を推定する



(c) Tomoyuki Furutani, Keio University SFC

55

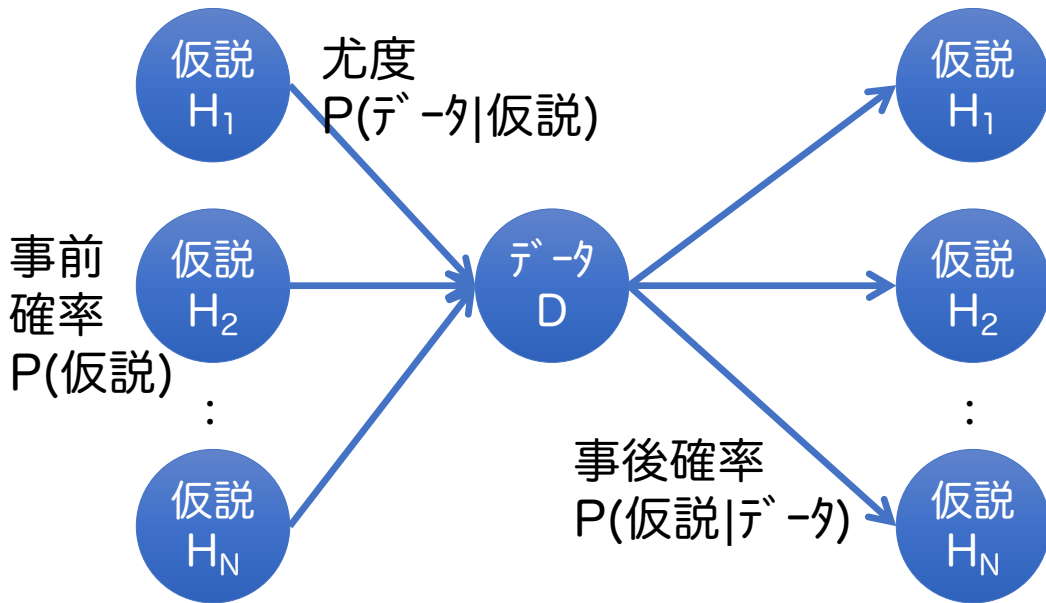
データから仮説（モデル）を推定する



(c) Tomoyuki Furutani, Keio University SFC

56

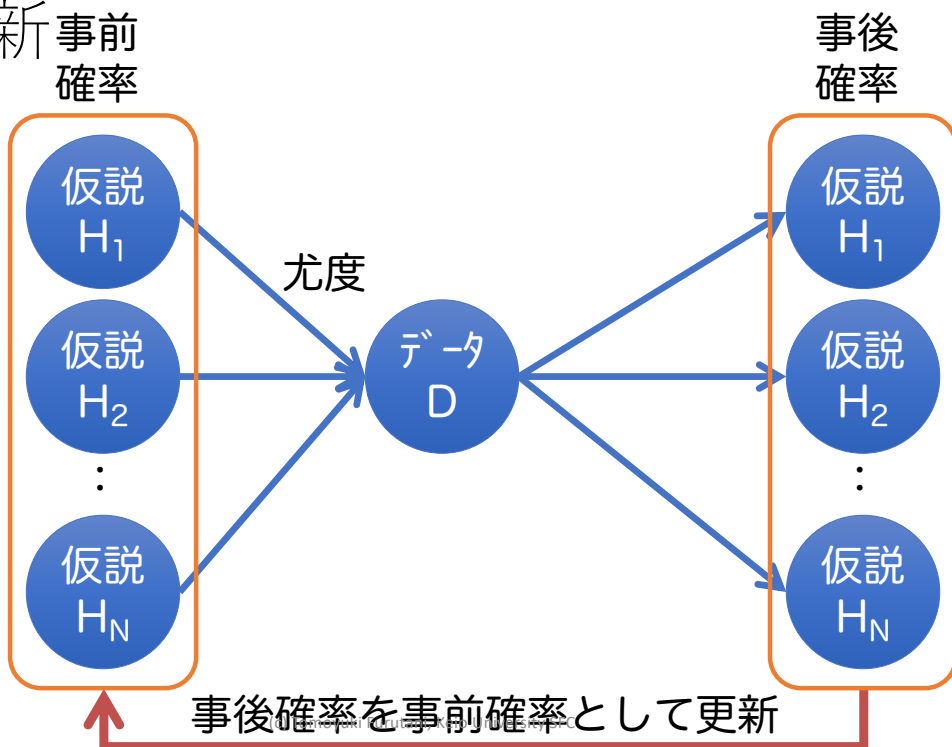
事前確率・尤度・事後確率



(c) Tomoyuki Furutani, Keio University SFC

57

ベイズ更新



58

ソフトウェア

- 授業の演習にはRというフリー統計ソフトを使います。
- 以下のサイトからダウンロード可能ですので各自インストールしてください。

<http://cran.r-project.org/>



CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

59

主要参考文献

- John K. Kruschke (著), 前田 和寛 (翻訳), 小杉 考司 (翻訳)(2017) 『ベイズ統計モデリング: R,JAGS, Stan によるチュートリアル』、共立出版
- 松浦 健太郎 (著), 石田 基広 (監修)(2016) 『StanとRでベイズ統計モデリング』、共立出版
- 久保 拓弥 (2012) 『データ解析のための統計モデリング入門』、岩波書店
- 安道 知寛 (2010) 『ベイズ統計モデリング』、朝倉書店
- 照井 伸彦 (2010) 『Rによるベイズ統計分析』、朝倉書店
- 涌井良幸 (2009) 『道具としてのベイズ統計』、日本実業出版社
- 古谷知之 (2008) 『ベイズ統計データ分析』、朝倉書店
- 松原望 (2008) 『入門ベイズ統計』、東京図書