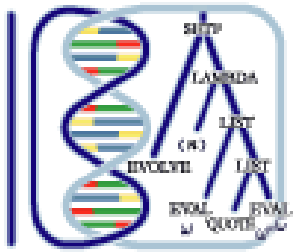
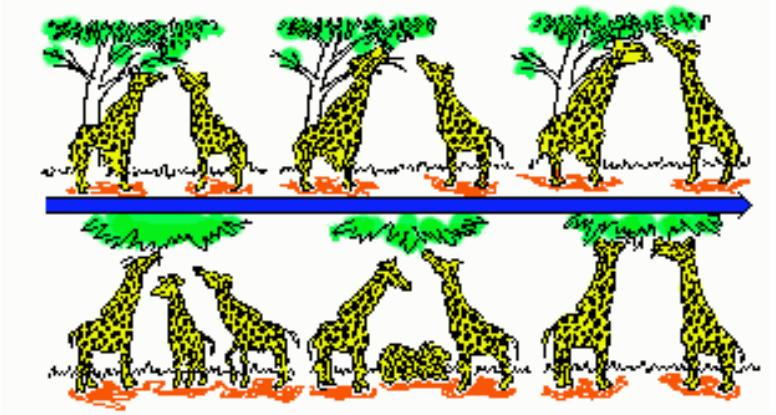


人工知能:最終回

東京大学大学院
情報理工学系科学研究科
電気情報学専攻
伊庭斉志



iba lab.

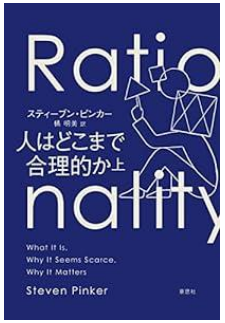
LLM の盲点：確率的なオウム

- 算数ができない
  - 9x9の演算はOK
  - 連立方程式？
- 論理的思考が苦手
  - 三段論法、背理法？
  - 経験でのみできるか？
- 生得的な知能はあるのか？
  - 人間などは進化の産物
  - 準備された学習
  - 遺伝子に組み込まれた学習の本能

では、人間はどうか？

ヒトと確率的推論

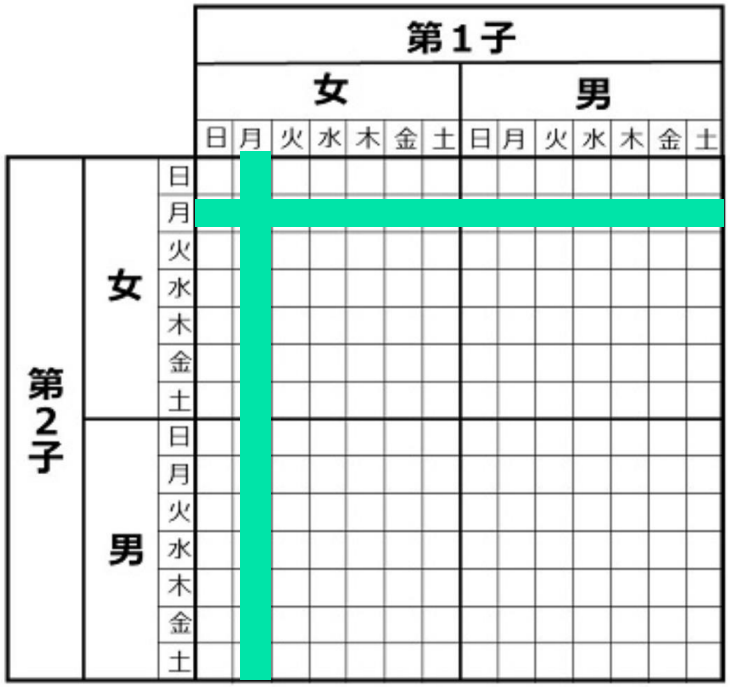
- 人も確率的推論が苦手である
- ランダム性の錯誤もしやすい
- スティーブン・ピンカー(1954-)
  - ハーバード大学の進化心理学者
  - 「人間は合理性に反して、ある事象に明確な原因があると断定することがある」



山田夫妻の子供：あらたな問い

質問3：山田夫妻には二人の子供がいる．そこで，「そのうちの一人は女の子だ」と言ったとする．このとき，もう一人の子供が女の子である確率をaとする（aは1/2ではない。aはいくつかを自分で考えてみよう）。では、「そのうちの一人は月曜日生まれの女の子だ」といったとする。その場合に、 もう一人の子供が女の子である確率はいくらか？　aと同じか異なるのかについて、根拠を示して説明せよ。

記述式テキスト（短文回答）



$$\frac{7+7-1}{14+14-1}=\frac{13}{27}$$

乳がんの確率

•あなたは乳ガンの検査（マンモグラフィ）で陽性になった。
•この検査では、乳ガンになっている人を見つける正確さは９０％、なっていない人を見つける正確さは９３％である。
•あなたが本当にガンである確率はいくつか？
•ただし、病気の発生確率は0.8％である。

•検査は90％近く正確なのだから、非常に可能性がある？？　（たいていの医者はこう答えた）

•本当か？

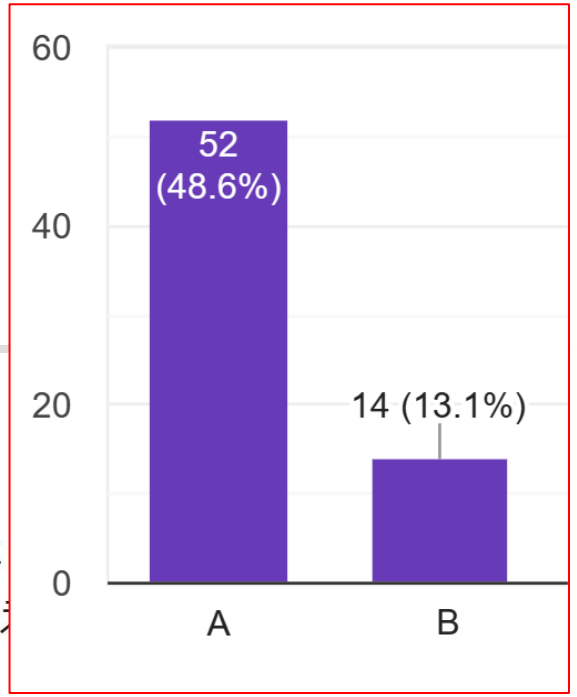
乳がんの確率

•女性が1000人いるとする。
•このうち８人がガンであり、かなり正確だが完ぺきではない検査によると、７人が陽性になる。
•あとの992人はガンではないが、検査はこの人々にとっても正確でない。
•992人のうち、70人近くの人にも陽性という結果が出てしまう。
•これらは偽陽性である。
•さて、全体では77人が陽性となるが、そのなかで確かに陽性なのは7人にすぎない。
•つまり、陽性になった女性のうち確かに陽性である可能性は１０％程度しかない。

ランダム性の錯誤

ンケートについて

の図は、一方が宇宙の星の並びを写
ものである。どちらが星の並びに見え



こちらが星座・ヒカリムシ

こちらがランダム

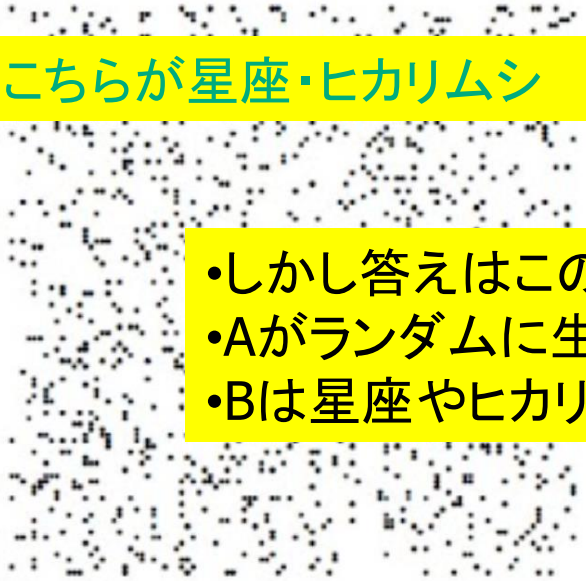


図 A

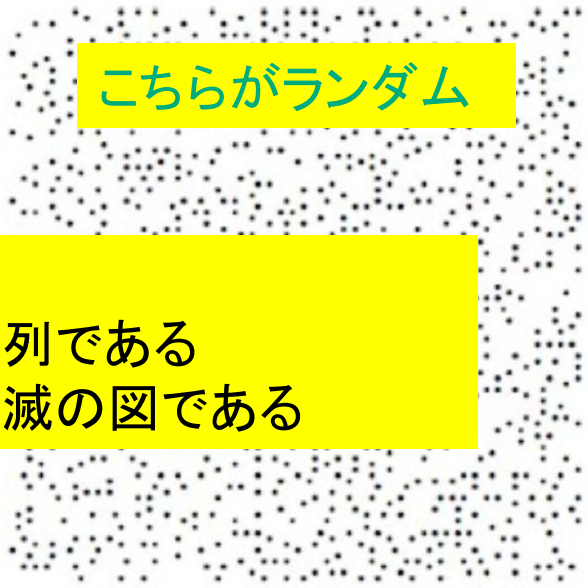


図 B

- しかし答えはこの逆
- Aがランダムに生成した点列である
- Bは星座やヒカリムシの点滅の図である

ランダム性の錯誤

- ポアソン分布の認識誤謬
  - 第二次大戦におけるロンドンへの爆撃
  - ギャンブラーの誤謬（ク
    1111111111と546253
  - ホットハンドの誤謬
    - 実はあった？？？

Surprised by the Gambler's and Hot Hand Fallacies?
A Truth in the Law of Small Numbers

Joshua B. Miller and Adam Sanjurjo *†‡

November, 2016 (with some updates)

Abstract

We prove that a subtle but substantial bias exists in a standard measure of the conditional dependence of present outcomes on streaks of past outcomes in sequential data. The magnitude of this novel form of selection bias generally decreases as the sequence gets longer, but increases in streak length, and remains substantial for a range of sequence lengths often used in empirical work. The bias has important implications for the literature that investigates incorrect beliefs in sequential decision making—most notably the Hot Hand Fallacy and the Gambler's Fallacy. Upon correcting for the bias, the conclusions of prominent studies in the hot hand fallacy literature are reversed. The bias also provides a novel structural explanation for how belief in the law of small numbers can persist in the face of experience.

JEL Classification Numbers: C12; C14; C18;C19; C91; D03; G02.
Keywords: Law of Small Numbers; Alternation Bias; Negative Recency Bias; Gambler's Fallacy; Hot Hand Fallacy; Hot Hand Effect; Sequential Decision Making; Sequential Data; Selection Bias; Finite Sample Bias; Small Sample Bias.

*Miller: Department of Decision Sciences and IGIER, Bocconi University, Sanjurjo: Fundamentos del Análisis Económico, Universidad de Alicante. Financial support from the Department of Decision Sciences at Bocconi University, and the Spanish Ministry of Economics and Competitiveness under project ECO2012-34928 is gratefully acknowledged.
†Both authors contributed equally, with names listed in alphabetical order.
‡This draft has benefitted from helpful comments and suggestions from the editor and anonymous reviewers, as well as Jason Abaluck, Jose Apesteguia, David Arathorn, Jeremy Arkes, Maya Bar-Hillel, Phil Birnbaum, Daniel Benjamin, Marco Bonetti, Colin Camerer, Juan Carrillo, Gary Charness, Ben Cohen, Vincent Crawford, Martin Dufwenberg, Jordan Ellenberg, Florian Ederer, Jonah Gabry, Andrew Gelman, Ben Gillen, Tom Gilovich, Maria Glymour, Uri Gneezy, Daniel Goldstein, Daniel Houser, Richard Jagacinski, Daniel Kahan, Daniel Kahneman, Erik Kimbrough, Dan Levin, Elliot Ludvig, Mark Machina, Daniel Martin, Filippo Massari, Guy Molyneux, Gidi Nave, Muriel Niederle, Christopher Olivola, Andreas Ortmann, Ryan Oprea, Carlos Oyarzun, Judea Pearl, David Rahman, Justin Rao, Alan Reifman, Pedro Rey-Biel, Yosef Rinott, Aldo Rustichini, Ricardo Serrano-Padial, Bill Sandholm, Vernon Smith, Lones Smith, Connan Snider, Joel Sobel, Charlie Sprenger, Daniel Stone, Sigrid Suetens, Dmitry Taubinsky, Richard Thaler, Nat Wilcox, and Bart Wilson. We would also like to thank seminar participants at Caltech, City U. London, Chapman U., Claremont Graduate School, Columbia U., Drexel U., George Mason U., New York University, NHH Norway, Max Planck Institute for Human Development (ABC), Microsoft Research, U. of

落雷の確率

質問2：あなたの住んでいる町では一年中いつでも落雷の可能性があるとする。その頻度は月に１回程度とする。つまり一日当たり落雷する確率は約0.03である。
さて、今日の月曜日にあなたの町に雷が落ちた。
では、次に落雷がある可能性が最も高いのは次のどれか？
理由をつけて答えよ。

(a)　明日の火曜日

(b)　一か月後

(c)　どの日も確率は変わらない

(d)　上のどれでも

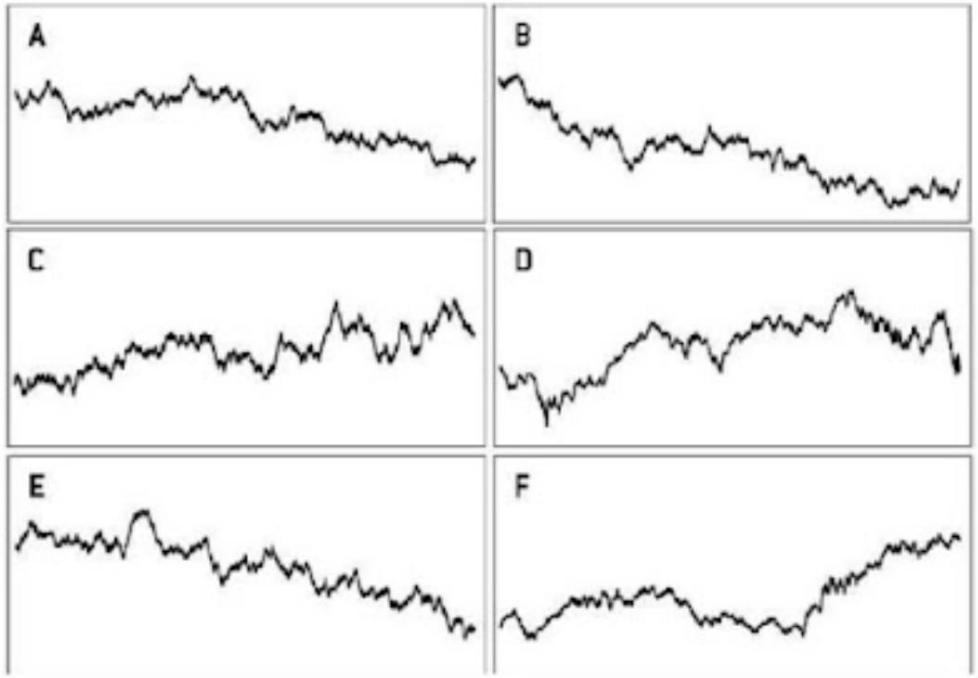
回答を入力

	落雷の確率 頻度　(回答)
正解	16.5%　(明日の火曜日)
不正解	67.4%　(どの日も確率は変わらない) 11.6%　(一か月後) 2.0%　(その他)

- 明日の火曜日：　確率0.03
- 明後日の水曜日：　0.97x0.03
- その次の木曜日：0.97x0.97x0.03
- 確率は指数関数的に減っていく

株の値動き

質問1：以下の６つの図のうちで２つは実際の株の値動きであり、残りの４つはランダムに生成したものである。株の値動きだと思われるの図を２つチェックせよ。



DとFは実際の株価チャート

Dは1970年代の株価推移
Fは1980年代最初の1000日間の株価変動

それ以外はコンピュータがランダムに生成した偽のチャート

テクニカル分析の難しさ!!

一億円のチャンス

無題のセクション

あなたは次のようなメールを有名な金持ちから受け取った。同じ手紙が20名に送られているという。メールは本物であり、次のように書かれている。『一億円を獲得するチャンスがあなたに与えられました。もし一億円が欲しいのなら、あなたの名前だけを書いてこのメールに返信してください。返信の期限は24時間で、ほかの誰からも返事が来ていなければ一億円はあなたのものです。まったく返事がなければ誰も賞金がもらえません。』このメールに対して、あなた（極めて合理的なエージェントとする）はどのような方針で返送するか・しないかを答えよ。ただし、自分以外の19人を探し出すなどの行動は禁止されている。

- 理性的で公平的なヒトであれば。。。
  - 1/20の確率で当たりのでるサイコロを振って当たれば返送する。
    はずれれば返送しない。

なぜそうしなかったのかを考えてみよう。
LLMではどうしただろうか？

強いAI v.s. 弱いAI

- 強いAI　　汎用型AI
- 弱いAI　　特化型AI

- 強いAIに必要なもの
  - 意識、「クオリア」を持つ
  - 身体性
- クオリアのなぞ
  - 逆転クオリア　　赤と緑が逆の人
  - 哲学的ゾンビ　　細胞まで同じだが全くクオリアがない、同じように笑い・泣く。

Why ChatGP???

Hans Peter Moravec (1948–)
アメリカの機械工学者、未来学者、トランスヒューマニスト。カーネギーメロン大学教授。

モラベックのパラドクス

- 人工知能は実現可能か？
- Easy
  - 数学的思考や論理的思考など
  - 専門的で高度な推論
- Hard
  - 幼児でも身に付けているような知覚
  - 運動の能力を獲得

- AIに仕事を奪われるのは？
  - ブルーカラー　v.s. ホワイトカラー

AI実現への批判：人類に代わるもの

- ユヴァル・ノア・ハラリ
  - アルゴリズムによる支配を予測
  - Chat-GPT???

- Future of Life Institute
  - 強力なAIシステムの開発を少なくとも6か月間停止する要求を発表
  - イーロン・マスク、スティーブ・ウォズアック、ヨシュア・ベンジオ
- ジェフリー・ヒントン
  - NNの２度の冬を越えた英雄
  - Deep LearningのGod Father
  - 2023年4月　Googleを退社
  - 「気兼ねなくAIの危険性を語るため」

Last words:...

伊庭研の研究テーマの1つ

- LLM for metaheuristics
  - Automatic program generation by LLM
  - GP augmented with LLM
- Metaheuristics for LLM
  - Genetic prompt engineering
  - Prompt learning by means of metaheuristics