

# 数理統計学の性質と AI の限界

—オープンソースカンファレンス 2024 Kyoto—

たいらようじ

2024-07-27

# 目次

自己紹介

現代の AI

数理統計学

数理統計学の性質

余談

AI の限界

日本の仕組み

AI の問題点

AI の活用

個人情報の問題

著作権の問題

AI と人

判断方法

問題点

総括

AI との関わり方

法律

今後の展望

# 自己紹介

## 発表者

たいらようじ

## 経歴

大学院修了 ( 修士 (理学) )。専門は代数幾何。

企業にてコンピュータの管理業務に従事。

高校で数学教員。

起業。

# 現代のAI

## AIの構成

AI = アルゴリズム + データ

アルゴリズム 数理統計学などを利用しデータを処理

データ AIの学習に利用

## AIの違い

アルゴリズムとデータの2つによりAIが決まる

# 数理統計学

## 数理統計学とは

**記述統計** 集めたデータの集団の性質などを調べる

**推測統計** 調べたい集団 (母集団) から標本として一部を取り出し、  
標本を調べることで元の集団を推測する

# 数理統計学

## 数理統計学とは

記述統計 集めたデータの集団の性質などを調べる

推測統計 調べたい集団 (母集団) から標本として一部を取り出し、  
標本を調べることで元の集団を推測する

- ▶ 推測統計は記述統計を基礎とし、確率論を用いて推測を行う
- ▶ 推測統計が数理統計学の主たる部分である
- ▶ 記述統計については 2012 年度高校入学生より必修となっっている

# 数理統計学

## 数理統計学とは

記述統計 集めたデータの集団の性質などを調べる

推測統計 調べたい集団 (母集団) から標本として一部を取り出し、  
標本を調べることで元の集団を推測する

- ▶ 推測統計は記述統計を基礎とし、確率論を用いて推測を行う
- ▶ 推測統計が数理統計学の主たる部分である
- ▶ 記述統計については 2012 年度高校入学生より必修となっっている

# 数理統計学

## 数理統計学とは

**記述統計** 集めたデータの集団の性質などを調べる

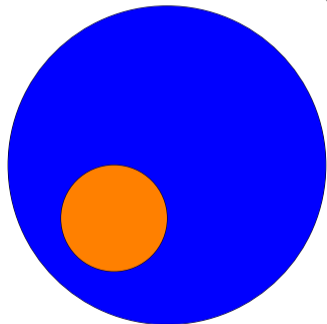
**推測統計** 調べたい集団 (母集団) から標本として一部を取り出し、  
標本を調べることで元の集団を推測する

- ▶ 推測統計は記述統計を基礎とし、確率論を用いて推測を行う
- ▶ 推測統計が数理統計学の主たる部分である
- ▶ 記述統計については 2012 年度高校入学生より必履修となっている



# 標本と母集団

母集団 (青) に対し、  
その一部分である標本 (オレンジ)

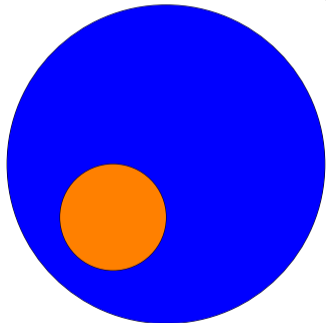


母集団から  $n$  個 ( $X_i$ ) 抽出する  
平均

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

# 標本と母集団

母集団 (青) に対し、  
その一部分である標本 (オレンジ)



母集団から  $n$  個 ( $X_i$ ) 抽出する  
平均

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

分散

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

不偏分散

$$\bar{X} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

# 回帰分析

気温とコーヒー売上

気温 ( $x$ )	5	7	8	15	19	23	27	28	25	20
杯数 ( $y$ )	48	63	80	96	73	98	130	114	83	81

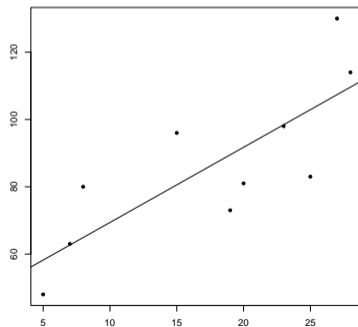
# 回帰分析

気温とコーヒー売上

気温 (x)	5	7	8	15	19	23	27	28	25	20
杯数 (y)	48	63	80	96	73	98	130	114	83	81

ある喫茶店のアイスコーヒーの売上と  
気温の回帰直線

$$y = 2.2349x + 47.0419$$



# 数理統計学の性質

## 重要な事実

数理統計学は因果関係を示さない

# 数理統計学の性質

## 重要な事実

数理統計学は因果関係を示さない

- ▶ 標本から得られる統計量から母集団の推測を行う
- ▶ 標本から得られる推測は標本が最もよく当てはまる性質から得ている
- ▶ 多数決のような方法で推測をしている為、イレギュラーを除いた中庸な結果が得られる

# 数理統計学の性質

## 重要な事実

数理統計学は因果関係を示さない

- ▶ 標本から得られる統計量から母集団の推測を行う
- ▶ 標本から得られる推測は標本が最もよく当てはまる性質から得ている
- ▶ 多数決のような方法で推測をしている為、イレギュラーを除いた中庸な結果が得られる

# 数理統計学の性質

## 重要な事実

数理統計学は因果関係を示さない

- ▶ 標本から得られる統計量から母集団の推測を行う
- ▶ 標本から得られる推測は標本が最もよく当てはまる性質から得ている
- ▶ 多数決のような方法で推測をしている為、イレギュラーを除いた中庸な結果が得られる



# 余談

## 帰納と演繹

### 統計と数学

「数理統計学は帰納」、「数学は演繹」である

**帰納** 個々の具体的な事例から一般に通用するような原理・法則などを導き出すこと

**演繹** 与えられた命題から、論理的形式に頼って推論を重ね、結論を導き出すこと

数学的帰納法とは帰納に似ている方法で行う演繹法のこと。

# 余談

## 帰納と演繹

### 統計と数学

「数理統計学は帰納」、「数学は演繹」である

**帰納** 個々の具体的な事例から一般に通用するような原理・法則などを導き出すこと

**演繹** 与えられた命題から、論理的形式に頼って推論を重ね、結論を導き出すこと

数学的帰納法とは帰納に似ている方法で行う演繹法のこと。

# AIの限界

## AIの構成

AI = アルゴリズム + データ

- ▶ アルゴリズム は、数理統計学 等を利用したデータ処理方法
- ▶ データ は、標本

## AIの性質

- ▶ 出力は確率的に起こりにくいものが除かれる
- ▶ データから得られる多数決のような結果 (近い値) を返す
- ▶ 学習データをどれほどうまく集めても必ず正しい答えを返す AI になるわけではない

# AIの限界

## AIの構成

AI = アルゴリズム + データ

- ▶ アルゴリズム は、数理統計学 等を利用したデータ処理方法
- ▶ データ は、標本

## AIの性質

- ▶ 出力は確率的に起こりにくいものが除かれる
- ▶ データから得られる多数決のような結果 (近い値) を返す
- ▶ 学習データをどれほどうまく集めても必ず正しい答えを返す AI になるわけではない

# AIの限界

## AIの構成

AI = アルゴリズム + データ

- ▶ アルゴリズム は、数理統計学 等を利用したデータ処理方法
- ▶ データ は、標本

## AIの性質

- ▶ 出力は確率的に起こりにくいものが除かれる
- ▶ データから得られる多数決のような結果 (近い値) を返す
- ▶ 学習データをどれほどうまく集めても必ず正しい答えを返す AI になるわけではない

# AIの限界

天動説と地動説

AIの出力は標本に多数ある似通ったデータに引きずられた結果となる

# AIの限界

## 天動説と地動説

AIの出力は標本に多数ある似通ったデータに引きずられた結果となる

### 天動説と地動説

コペルニクス以前にデータを収集すれば  
AIは天動説が正しいという結果を示し、  
以後であれば地動説が正しいという結果を示す

# AIの限界

## 天動説と地動説

AIの出力は標本に多数ある似通ったデータに引きずられた結果となる

### 天動説と地動説

コペルニクス以前にデータを収集すれば  
AIは天動説が正しいという結果を示し、  
以後であれば地動説が正しいという結果を示す

AIは多数決に似たような処理であるので  
多数決で決定してよい状況で利用するのに適している



# AIの限界

## 2つの分布

学習データが異なる分布のデータが混じっている場合、意図しない結果が得られる可能性がある

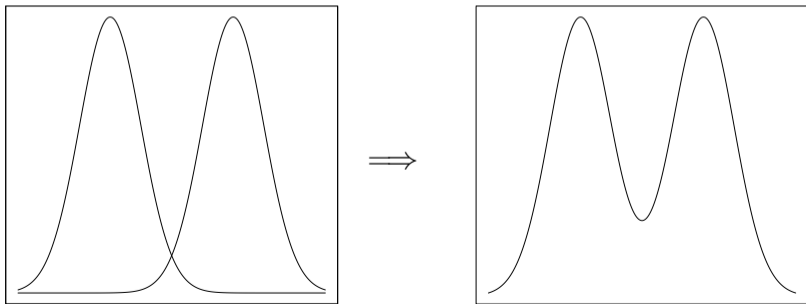
# AIの限界

## 2つの分布

学習データが異なる分布のデータが混じっている場合、意図しない結果が得られる可能性がある

## 雌雄の差

ある魚の平均体長を求めるとき、雌雄で大きく体長の異なるなら求めた平均体長の個体は存在しない可能性がある。



# AIの限界

近い値

AIにより求めた値が欲しかったものと少々異なっても良い場合はAIを用いても影響は少ない

# AIの限界

近い値

AIにより求めた値が欲しかったものと少々異なっても良い場合はAIを用いても影響は少ない

## 誤差の許容

「 $2 + 3$ 」の結果をAIに答えさせた時、 $2 + 3 = 4$  より  $2 + 3 = 4.99$  とする答えが良い結果であるとする場合はAIを使うメリットが有る

$2 + 3 = 5$  以外の答えには意味がない場合はAIを使うメリットは無い

# 日本の仕組み

## 日本

- 民主主義 一人ひとりが権利 (社会権、自由権) をもつ政治の形態  
社会権のほうが自由権より強い
- 資本主義 資本 (お金) を自由に稼ぐことが出来る経済の形態

# 日本の仕組み

## 日本

**民主主義** 一人ひとりが権利 (社会権、自由権) をもつ政治の形態  
社会権のほうが自由権より強い

**資本主義** 資本 (お金) を自由に稼ぐことが出来る経済の形態

## 多数決について

多数決は資本主義の決定方法であり民主主義の決定方法ではない  
民主主義の理想的な決定方法は全会一致であり、少数の意見が切り捨てられるのは民主主義に反する

# AIの問題点

## AIの活用

### AIの利用場面

- ▶ 資本主義の決定方法が利用可能なとき AIは大きな力を発揮する
- ▶ 人の権利にかかわるような場合や民主主義の決定方法が必要な場面では AIに判断を委ねてはいけない

# AIの問題点

## AIの活用

### AIの利用場面

- ▶ 資本主義の決定方法が利用可能なとき AIは大きな力を発揮する
- ▶ 人の権利にかかわるような場合や民主主義の決定方法が必要な場面では AIに判断を委ねてはいけない



# AIの問題点

## AIの活用

### AIの利用場面

- ▶ 資本主義の決定方法が利用可能なとき AIは大きな力を発揮する
- ▶ 人の権利にかかわるような場合や民主主義の決定方法が必要な場面では AIに判断を委ねてはいけない

### リクナビ事件(2019年)

学生の情報を収集し内定辞退率をAIにて算出し企業へ販売  
個人情報保護委員会、厚労省等により指導勧告

# AIの問題点

## AIの活用

### AIの利用場面

- ▶ 資本主義の決定方法が利用可能なとき AIは大きな力を発揮する
- ▶ 人の権利にかかわるような場合や民主主義の決定方法が必要な場面では AIに判断を委ねてはいけない

### リクナビ事件(2019年)

学生の情報を収集し内定辞退率をAIにて算出し企業へ販売  
個人情報保護委員会、厚労省等により指導勧告

### 三重県女児虐待死事件(2023年)

児童相談所にAIを活用したシステム導入  
システムの判定が39%であった為、一時保護を見送り

# AIの問題点

## 個人情報の問題

### 学習データに集める標本

- ▶ 個人情報の収集方法  
同意の有無や利用方法の説明
- ▶ 個人データの管理  
漏洩や外部への提供への制限等
- ▶ 個人情報保護法  
第一条  
(略) 個人の権利利益を保護することを目的とする

# AIの問題点

## 著作権の問題

AIは学習データから出力結果を得ている

学習データが著作権を持つものを含んでいる場合、出力結果も元の著作権が含まれる可能性がある

## 著作権法

第二条 第1項 第一号

著作物 思想又は感情を創作的に表現したものであつて、文芸、学術、美術又は音楽の範囲に属するものをいう。

.....  
第三十条の四

著作物は、次に掲げる場合(略)利用することができる。

- 一 著作物の録音、録画その他の利用に係る技術の開発又は実用化のための試験の用に供する場合
- 二 情報解析(略)の用に供する場合
- 三 (略)当該著作物を電子計算機による情報処理の過程における利用その他の利用(略)に供する場合

# AIと人

## 判断方法

### 人間が判断を行うときの材料

1. 感情的 (感情や本能、欲求など)
2. 統計的 (過去の体験や他人の意見など)
3. 論理的

上記の3つを複合的に利用している

# AIと人

## 判断方法

### 人間が判断を行うときの材料

1. 感情的 (感情や本能、欲求など)
2. 統計的 (過去の体験や他人の意見など)
3. 論理的

上記の3つを複合的に利用している

### AIの判断方法

AIは2 統計的な判断を行う

3 論理的な判断 は数理統計学を利用した AI では実現不可能

# AIと人

## 判断方法

### 人間が判断を行うときの材料

1. 感情的 (感情や本能、欲求など)
2. 統計的 (過去の体験や他人の意見など)
3. 論理的

上記の3つを複合的に利用している

### AIの判断方法

AIは2 統計的な判断を行う

3 論理的な判断 は数理統計学を利用したAIでは実現不可能

学術では2 統計 と3 論理 を併用することが多い

# AIと人

## 問題点

### 学習の違い

- ▶ 人：統計的な情報から理論を理解し他への応用を行う
- ▶ AI：統計的な情報から推測を行う



# AIと人

## 問題点

### 学習の違い

- ▶ 人：統計的な情報から理論を理解し他への応用を行う
- ▶ AI：統計的な情報から推測を行う

### 問題点

- ▶ 人も統計的な判断を行う為、AIを人のように錯覚する
- ▶ 統計的な判断は正しい答えに辿り着く道筋を示せない
- ▶ 論理的な判断は道筋を示すことが出来る為、法的な判断などで必要

# AIと人

## 問題点

### 学習の違い

- ▶ 人：統計的な情報から理論を理解し他への応用を行う
- ▶ AI：統計的な情報から推測を行う

### 問題点

- ▶ 人も統計的な判断を行う為、AIを人のように錯覚する
- ▶ 統計的な判断は正しい答えに辿り着く道筋を示せない
- ▶ 論理的な判断は道筋を示すことが出来る為、法的な判断などで必要

# AIと人

## 問題点

### 学習の違い

- ▶ 人：統計的な情報から理論を理解し他への応用を行う
- ▶ AI：統計的な情報から推測を行う

### 問題点

- ▶ 人も統計的な判断を行う為、AIを人のように錯覚する
- ▶ 統計的な判断は正しい答えに辿り着く道筋を示せない
- ▶ 論理的な判断は道筋を示すことが出来る為、法的な判断などで必要

# 総括

## AIとの関わり方 AIの性質

- ▶ AIは中庸な結果を返すだけで正しい答え返していない
- ▶ AIの出力は求めたい値に近い値を返している
- ▶ AIの出力の妥当性は人が判断しないとイケない

# 総括

## AIとの関わり方 AIの性質

- ▶ AIは中庸な結果を返すだけで正しい答え返していない
- ▶ AIの出力は求めたい値に近い値を返している
- ▶ AIの出力の妥当性は人が判断しないとイケない

# 総括

## AIとの関わり方 AIの性質

- ▶ AIは中庸な結果を返すだけで正しい答え返していない
- ▶ AIの出力は求めたい値に近い値を返している
- ▶ AIの出力の妥当性は人が判断しないとイケない

# 総括

## AIとの関わり方 AIの性質

- ▶ AIは中庸な結果を返すだけで正しい答え返していない
- ▶ AIの出力は求めたい値に近い値を返している
- ▶ AIの出力の妥当性は人が判断しないとイケない

## 人が身につけておくべきこと

- ▶ 論理的思考能力
- ▶ 数理統計学の知識

# 総括

## AIとの関わり方 AIの性質

- ▶ AIは中庸な結果を返すだけで正しい答え返していない
- ▶ AIの出力は求めたい値に近い値を返している
- ▶ AIの出力の妥当性は人が判断しないとイケない

## 人が身につけておくべきこと

- ▶ 論理的思考能力
- ▶ 数理統計学の知識
  
- ▶ 現在の中等教育では論理的な思考能力はほとんど身につかない
- ▶ 人は論理的思考を身につけておくべきである
- ▶ ユリウス カエサル **Julius Caesar** 「人は見たいものしか見ない」



# 総括

## AIとの関わり方 AIの性質

- ▶ AIは中庸な結果を返すだけで正しい答え返していない
- ▶ AIの出力は求めたい値に近い値を返している
- ▶ AIの出力の妥当性は人が判断しないとイケない

## 人が身につけておくべきこと

- ▶ 論理的思考能力
- ▶ 数理統計学の知識
  
- ▶ 現在の中等教育では論理的な思考能力はほとんど身につかない
- ▶ **人は論理的思考を身につけておくべきである**  
ユリウス カエサル
- ▶ **Julius Caesar**「人は見たいものしか見ない」

# 総括

## AIとの関わり方 AIの性質

- ▶ AIは中庸な結果を返すだけで正しい答え返していない
- ▶ AIの出力は求めたい値に近い値を返している
- ▶ AIの出力の妥当性は人が判断しないとイケない

## 人が身につけておくべきこと

- ▶ 論理的思考能力
- ▶ 数理統計学の知識
  
- ▶ 現在の中等教育では論理的な思考能力はほとんど身につかない
- ▶ 人は論理的思考を身につけておくべきである
- ▶ ユリウス カエサル **Julius Caesar** 「人は見たいものしか見ない」

- ▶ 民主主義に関わる判断を AI に委ねるのには問題がある
- ▶ 学習データに個人情報を含める場合、個人情報保護法による制約がある
- ▶ 学習データに著作権があれば AI の出力は著作権上の問題を含む可能性がある
- ▶ 数理統計学の理解が無ければ AI をうまく利用することが難しい

# 総括

## 法律

- ▶ 民主主義に関わる判断を AI に委ねるのには問題がある
- ▶ 学習データに個人情報を含める場合、個人情報保護法による制約がある
- ▶ 学習データに著作権があれば AI の出力は著作権上の問題を含む可能性がある
- ▶ 数理統計学の理解が無ければ AI をうまく利用することが難しい

# 総括

## 法律

- ▶ 民主主義に関わる判断を AI に委ねるのには問題がある
- ▶ 学習データに個人情報を含める場合、個人情報保護法による制約がある
- ▶ 学習データに著作権があれば AI の出力は著作権上の問題を含む可能性がある
- ▶ 数理統計学の理解が無ければ AI をうまく利用することが難しい

# 総括

## 法律

- ▶ 民主主義に関わる判断を AI に委ねるのには問題がある
- ▶ 学習データに個人情報を含める場合、個人情報保護法による制約がある
- ▶ 学習データに著作権があれば AI の出力は著作権上の問題を含む可能性がある
- ▶ 数理統計学の理解が無ければ AI をうまく利用することが難しい

# 今後の展望

AI をうまく利用できるのか？

- ▶ **教育分野への利用**
- ▶ 人材教育 (数理統計学の理解、論理的思考能力の強化)
- ▶ 人材確保 (開発者、利用者)
- ▶ 新技術への発展

# 今後の展望

AI をうまく利用できるのか？

- ▶ 教育分野への利用
- ▶ 人材教育 (数理統計学の理解、論理的思考能力の強化)
- ▶ 人材確保 (開発者、利用者)
- ▶ 新技術への発展



# 今後の展望

AI をうまく利用できるのか？

- ▶ 教育分野への利用
- ▶ 人材教育 (数理統計学の理解、論理的思考能力の強化)
- ▶ 人材確保 (開発者、利用者)
- ▶ 新技術への発展

# 今後の展望

AI をうまく利用できるのか？

- ▶ 教育分野への利用
- ▶ 人材教育 (数理統計学の理解、論理的思考能力の強化)
- ▶ 人材確保 (開発者、利用者)
- ▶ 新技術への発展