# ミニレクチャー 「統計数理」の研究について

統計数理研究所 (略称: 統数研) 2020年10月27日 オープンハウス



第2代および5代 末綱恕一所長による筆耕 統数研の象徴「数」のレリーフ @立川総合研究棟アトリウム



### 統計学とはどんな学問なのか

**K.Pearson**;統計学を現象の観察に基づいて科学的法則を導く「**科学の文法**」として構想ありとあらゆる対象が統計的認識プロセスを通じて科学となる

C. R, Rao「統計学とは何かー偶然を活かす」:統計学は,不確実性を数量化することで, 自然や社会にあふれる偶然(未知の動き)に立ち向かう科学.

Stanford Encyclopedia of Philosophy: 殆ど全ての分野の科学的発見をサポート

現代:データの価格崩壊時代→科学だけでなく産業や社会の意思決定もサポート

実験・調査の最適計画

統計的問題解決

データの最適収集・構造化

データへの適用 「データ解析」 現象のモニタリング 統計的問題発見 問題の原因を究明 問題解決法を提示 モデル評価, モデル選択, モデルの改善

統計モデル統計的決定理論

モデリング理論

予測・制御 因果効果予測



統計科学的意思決定とその社会実装

統計的問題解決の貢献

科学的仮説の検証:自然科学・社会科学の法則最適行動の選択:新たな治療法の統計的決定

最適なオペレーションの提示

### 統計学の歴史

#### 国勢学と古典確率論

- 1. 国勢学派(ドイツ) Statistik Achenwall(1712-1772)
   国家の顕著事象の記録
- 2.政治算術学派 (イギリス)Graunt (1620-1674)社会現象の数量的観察,規則性の指摘,推算
- 3. 古典確率論 (フランス)
  - 16世紀 Cardano
  - 17世紀 Fermat, Pascal, Bernoulli「推測法」
  - 18世紀 de Moivre「偶然論」, Bayes
  - 19世紀 Laplace「確率の解析的理論」

#### 数理と統計の合体

Gauss (1777-1855): : 誤差論・最小二乗法 1830-1850 官庁統計整備・社会の計測 Quetelet (1796-1874): 数理天文学者 大量観測による法則定立と統計に基づく社会改善

#### 20世紀:近代統計学の誕生と進化

1. 「科学の文法」としての統計学

Galton: 四分位数,相関係数,回帰係数

K. Pearson:科学の文法

記述統計学的方法(ヒストグラム)

統計モデルの適合度

20世紀初頭 計量諸科学の創成

2. 推測統計学(母集団についての仮説検証)

Gosset: t-分布の発見と小標本理論

Fisher: 尤度原理と実験計画法

Neyman: 仮説検定・区間推定と標本調査論

Shewhart · Deming: PDCAサイクルと品質管理

Wald: 統計的決定理論と逐次推論

3. データの科学・データからの仮説発見

Tukey (1962) データ解析

統数研:現象と行動の統計数理 + 予測統計学

# 統数研の歴史

#### 設置目的 「統計数理に関する総合研究」

●1944年: 文部省直轄の研究所として設立

●1985年:大学共同利用機関に改組転換

●1988年:総合研究大学院大学創設

●2004年:法人化,機構化

●2009年:立川移転

設立

●2010年:Akaike Guest House設立

●2014年:URA(University Research Administrator)を配置

●2019年:6月5日 75周年記念式典

大学共同利用機関

大学院

#### 主要活動

#### 〇 研究活動

- ・我が国における統計数理の中核拠点
- · NOE(Network Of Excellence)形成事業
- ・先端的な研究を推進

#### 〇 共同利用

- ・多様な分野との共同研究
- ・スパコン、ソフトウェア、乱数

#### 〇 人材育成

- ・総研大における大学院教育
- ·統計思考力育成事業

情報・システム 研究機構 <u>立川移</u>転

創立75周年

Akaike Guest House 設立

1944 19851988 2004 2009 2010 2014 2019

URA配置

### 「統計数理研究所のこれまでの研究」とは

- ◆ 複 雑 ・ 不 確 実 ・ 動 的 な 対 象
- ◆ 不 完 全 情 報 下 で の 最 適 意 思 決 定

統計数理とは:実世界の現象を解明するために,本質的な情報を効率的に抽出し,構造化することで<u>予測・知識</u>獲得や最適なオペレーションを行うための<u>方法</u>を対象とする横断的科学



- ・調査・実験計画
- 統計モデル
- ·統計的推論
- ·統計的予測
- ・問題解決プロセス

- •調査の科学
- ●モデル選択
- •ベイズモデリング
- •粒子フィルタ
- ●データ同化
- •スパースモデリング
- •カーネル法等々



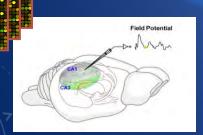
データ環境の変化とともに変容





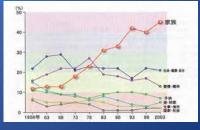
ビッグデータ. クラウド

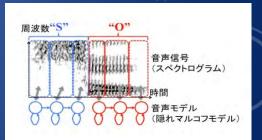


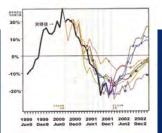












# 「統計数理」の特性

### 統計数理は

- データに基づく科学の方法 鍵となるのは「データ」,「モデル」,「推論」
- 社会からの要請等に応じて変わる 変わっていく中での<u>基幹</u>となる 普遍的考え方や方向性が存在

様々な分野を<u>横断し、つなぐ「モデル」</u>、そして様々な分野を<u>横断し、考えを巡らす「推論」</u>、という特性、いわば

データに基づく科学の生成様式とその支援ツール

を研究する横断的基幹学問領域

### 研究組織の変遷(1)

1944年

所長

所員

助手

書記

1946年

#### 第一部 (基礎理論)

- 1. 一般確率論・解析的確率論
- 2. 数理統計論
- 3. 応用解析論・数学解析論・関数論
- 4. 確率論的力学・力学・応用幾何学

1949年

#### 研究第一部

確率論 数理統計の基礎理論 確率統計に関する基礎研究

第二部(自然科学)

- 1. 応用統計論
- 2. 図表学
- 3. 数值解析方法論·誤差数值計算
- 4. 調和解析論・周期解析論

研究第二部

数理統計学の理論・自然科学 方面への応用 実験統計 1955年

#### 第一研究部

- 1. 一般統計推論
- 2. 過程事象の解析
- 3. 統計推論の基礎となる確率論

第二研究部

- 1. 標本調査法・社会調査法
- 2. 統計に関する実験の計画法, 測定法
- 3. 一般統計事象の分析法・管理法
- 4. 一般現象を数量的に取り扱う方法
  - ·現象予測法

#### 第三部(社会科学)

- 1. 資料論
- 2. 実験·統計的実験
- 3. 計器研究·計器測定

#### 研究第三部

数理統計学の理論・社会現象 への応用 計量統計 統計計算

#### 第三研究部

- 1. 統計解析に必要な計算法
- 2. 統計解析の機械化・統計計算に 必要な計算機
- 3. 企画管理についての数理的方法 研究指導普及室

#### 1971年

#### 第一研究部

- 1. 一般統計推論
- 2. 過程事象の解析
- 3. 統計推論の基礎となる確率論

#### 第二研究部

- 1. 標本調査法・社会調査法
- 2. 統計に関する実験の計画法, 測定法
- 3. 現象を数量的に取り扱う方法及び 統計的解析法

#### 第三研究部

- 1. 企画管理の数理的方法
- 2. 過程管理の数理的方法 研究指導普及室

#### 第四研究部

- 1. 数値解析と計算機システムの研究
- 2. 統計的パターン認識と統計的シミュレーション

#### 第五研究部

- 1. 予測理論に関する研究
- 2. 制御理論に関する研究

#### 第六研究部

1. 行動に関する統計的計量法及びデータ解析法

#### 1985年

#### 統計基礎研究系

推測決定理論研究部門 基礎概念研究部門 応用確率論研究部門 確率・分布理論研究部門

#### 調查実験解析研究系

標本調査研究部門 空間事象研究部門 多次元解析研究部門 パターン解析研究部門 系列事象研究部門 実験計画研究部門

#### 予測制御研究系

システム解析研究部門 予測理論研究部門 制御理論研究部門 数値的最適化研究部門 統計計算システム研究部門 非数値情報処理研究部門 大規模システム研究部門

#### 領域統計研究系

自然科学領域研究部門 人文社会科学研究部門 複合領域研究部門

# 研究組織の変遷2

1988年

#### 総合研究大学院大学創設





#### データから有効な予測や新知見を引き出 す

複雑な対象、氾濫する情報と不確実性のもと、データに よる合理的な推論、有効な予測、新知見の発見のための 統計モデルや統計的方法を研究するのが統計…

- 統計科学専攻 □
- 統計数理研究所 ②

# 研究組織の変遷3

大学共同利用 機関法人化

所長

副所長 (兼)

研究主幹 (兼)

教授

准教授

助教

2004年

#### 統計基礎研究系

推測決定理論決定 基礎概念研究 応用確率論研究 確率・分布理論研究(客員)

#### 調查実験解析研究系

標本調査研究 空間事象研究 多次元解析研究 パターン解析研究 系列事象研究 実験計画研究(客員)

#### 予測制御研究系

システム解析研究 予測理論研究 制御理論研究 数値的最適化研究 統計計算システム研究 非数値的情報処理研究 大規模システム研究(客員)

#### 領域統計研究系

自然科学領域研究 人文社会科学領域研究 複合領域研究(客員) 2005年

モデリング研究系

時空間モデリング

知的情報モデリング

グラフ構造モデリング

#### 2012年

#### モデリング研究系

時空間モデリング <u>複雑構造モデリング</u> 潜在構造モデリング 2018年~現在

#### モデリング研究系

予測制御 複雑構造モデリング データ同化

#### データ科学研究系

調査解析 多次元データ解析 計算機統計

#### データ科学研究系

データ設計 <u>計量科学</u> 構造探索

#### データ科学研究系

調査科学 計量科学 構造探索

#### 数理・推論研究系

統計基礎数理 学習推論 計算数理

#### 数理・推論研究系

統計基礎数理 学習推論 計算推論

#### 数理・推論研究系

統計基礎数理 学習推論 数理最適化

# 統数研の主要3事業

高度 統計科学技術センター NOE型研究人材育成 (計算研究基盤) 統計数理の適用に 統計 関わる重点分野共同研究 に関わる数理的研究 基 数理·推論研究系 統 計思考院 幹 モデリング研究系 的 研 データ科学研究系 究

# 現在の統数研の研究組織

モデリング研究系 Department of Statistical Modeling

データ科学研究系 Department of Statistical Data Science

数理·推論研究系
Department of Statistical Inference
and Mathematics

-予測制御グループ
- 複雑構造モデリンググループ
- 複雑構造モデリンググループ
- データ同化グループ
- 調査科学グループ
- 構造探索グループ
- 横計基礎数理グループ
- 学習推論グループ
- 数理最適化グループ

基幹的研究組織

リスク解析戦略研究センター

統計的機械学習研究センター

ものづくりデータ科学研究センター

医療健康データ科学研究センター

人材育成組織 統計思考院

統計科学技術センター

図書室

NOE型研究組織

### 統数研の組織運営体制

■所長 椿 広計 (平成31.4.1-第1期2年目)

■副所長 伊藤 聡 (研究企画・人事) 山下智志 (財務·知財) 宮里 義彦(評価・広報)







山下智志

■研究主幹

松井知子(モデリング研究系) 金藤浩司(データ科学研究系) **栗木** 哲(数理·推論研究系)







■統計科学技術センター長

上野玄太 副センター長 南 和宏





○統計思考院長 川崎能典 副院長 伊庭幸人





伊庭幸人

▼総合研究大学院大学 複合科学研究科 統計科学専攻長

藤澤洋徳



藤澤洋徳

■ NOE型研究センター長

山下智志 (リスク解析戦略研究センター)

福水健次 (統計的機械学習研究センター)

吉田 亮 (ものづくりデータ科学研究センター)

松井茂之(医療健康データ科学研究センター)





# 「統計数理」を研究する3つの系

〇モデリング研究系

〇データ科学研究系

〇数理·推論研究系

### 統数研の革新的な研究成果

情報量規準 統計モデルの良さを客観的に比較できる

数量化理論 感性などの質的なデータも解析できる

時系列解析 時間とともに変わる複雑な現象を予測・制御できる

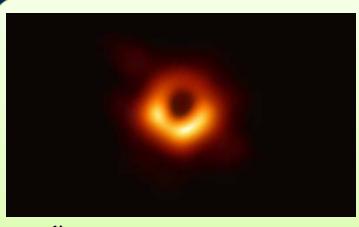
ベイズモデリング 潜在的な構造や因子を浮き彫りにできる

データ同化 データを使ってシミュレーションを系統的に改良できる

統計数理研究所 創立75周年 プレスリリースより一部抜粋 https://www.ism.ac.jp/ura/press/ISM2018-01.html

Alを実現するための理論や方法ともなる 統計的機械学習

# 統数研のNOE型研究(共同研究)の成果等



credit: Event Horizon Telescope Collaboration 方法が使われた。

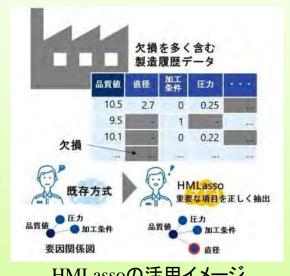
Event Horizon Telescope (EHT)の最初の研究成果と して発表された超巨大質量 ブラックホールシャドウの 画像。スパースモデリング と呼ばれるデータ科学の

大量の欠損を含むデータからでも不 具合の要因を特定する機械学習アル ゴリズム

「HMLasso (Least absolute shrinkage and selection operator with High Missing rate) を(株)東芝と共同研究で開発した。 品質低下や歩留悪化などの要因を高 速・高精度に特定(最先端のアルゴ

リズム「CoCoLasso」と比べて推定

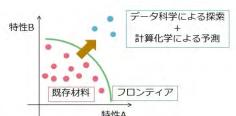
誤差を約41%削減することに成功)



HMLassoの活用イメージ

#### 民間と統数研で共同件研究部門の設置

#### 1)三菱ケミカル株式会社



本研究部門で構築するアルゴリズムは学術成果として 積極的に発信。マテリアルズ・インフォマティクス分 野におけるオープンイノベーション・オープンサイエ ンスの促進に貢献

#### 2)JSR株式会社



多次元にわたる材料空間の新しい可視化や表現手法の開 発、高速な計算化学手法を用いたデータ蓄積による性能 予測モデルの適用範囲の拡大、理論的な洞察に基づいた 新しい物性記述子の開発など化学とデータ科学を融合さ せたアルゴリズムを開発しソフトウェアとして実装

# 「統計数理」を研究する仲間たち

総合研究棟1階アトリウム「数」レリーフ前にて撮影

