

CENTER FOR DATA SCIENCE

データサイエンス+α

—意思決定画像を用いたデータ科学体系化への試み—



早稲田大学 データ科学教育チーム
小林, 野村, 須子, 安田, 堀井, 鎌塚, 中原, 松嶋



2017年頃からの各大学におけるデータサイエンス教育がスタート

各大学の動き

文科省 : 「数理及びデータサイエンスに係る教育強化」の拠点校（6校）

北海道大学 : 数理・データサイエンス教育研究センター

東京大学 : 数理・情報教育研究センター

京都大学 : データ科学イノベーション教育研究センター

大阪大学 : 数理・データ科学教育研究センター

九州大学 : 数理・データサイエンス教育研究センター

滋賀大学 : データサイエンス教育研究センター
データサイエンス学部

データサイエンス教育の全学・全国への展開

横浜市立大学 : データサイエンス学部

武蔵野大学 : データサイエンス学部

群馬大学 : 数理データ科学教育研究センター

全国の大学も
データサイエンス
のセンターや
学部の設立を
進めている

早稲田大学データ科学センター は2017年発足2018年本格稼働



CDS 早稲田大学
データ科学センター

お問い合わせ先

【ホームページ】

<https://www.waseda.jp/inst/cds/>



【電話】

03-3204-9164

【E-mail】

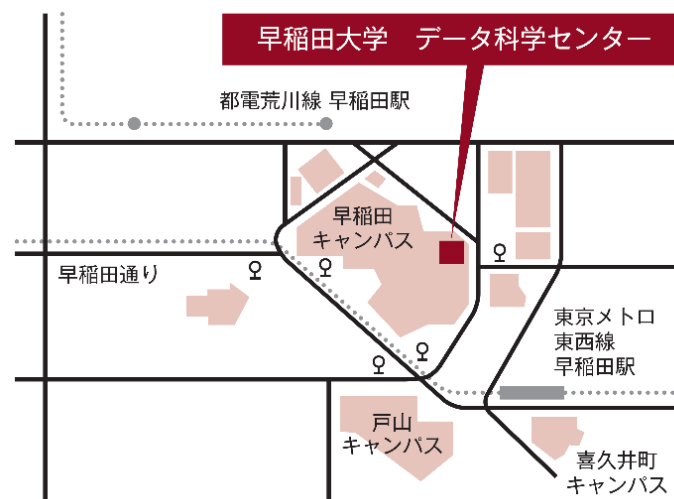
cds-info@list.waseda.jp

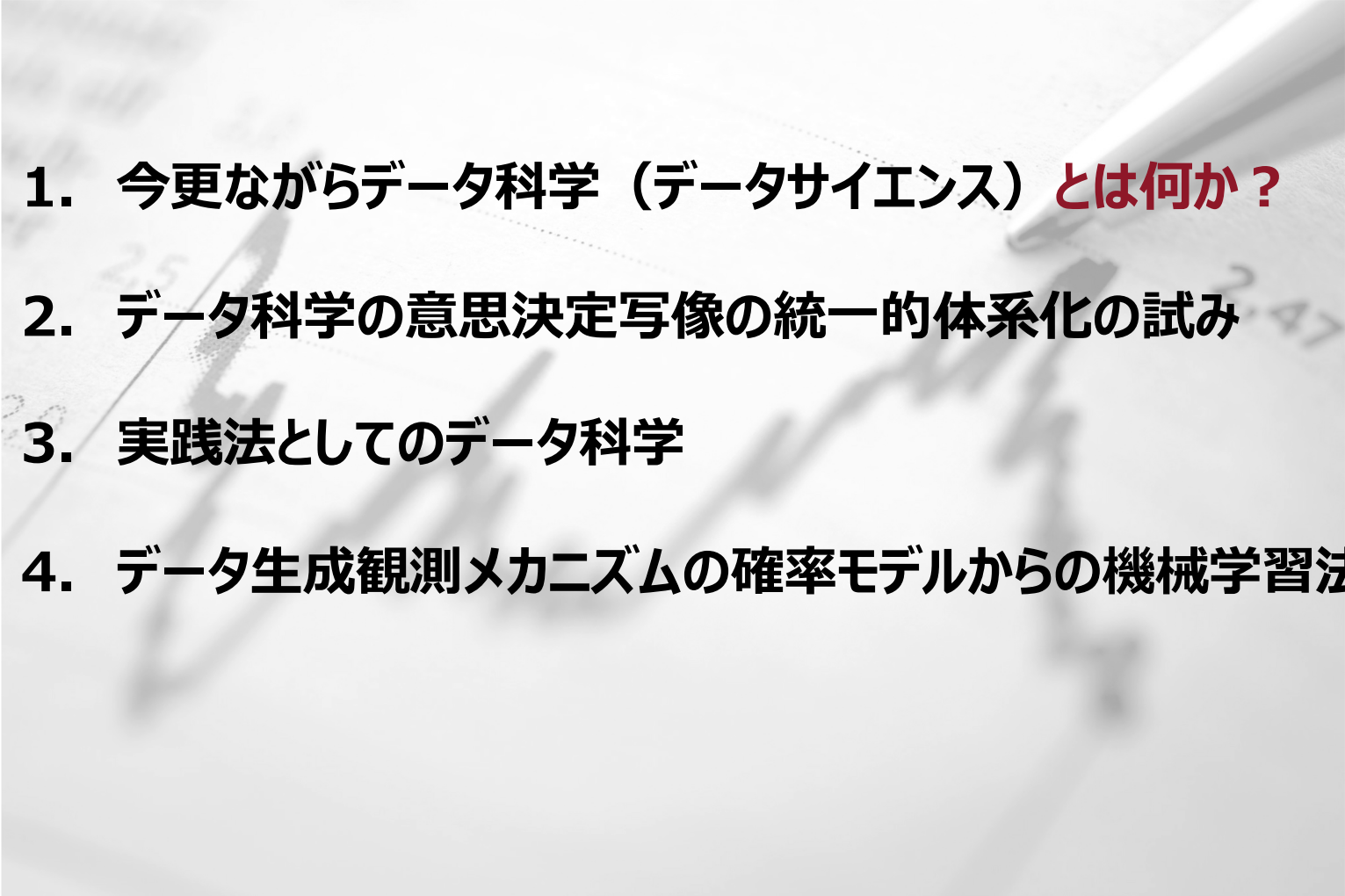
【担当】

梅原事務長、田尻、草野、松木、
高橋、箱川

交通アクセス

169-8050 東京都新宿区西早稲田1-6-1 1号館3階
東京メトロ東西線 早稲田駅 徒歩5分



- 
1. 今更ながらデータ科学（データサイエンス）とは何か？
 2. データ科学の意思決定写像の統一的体系化の試み
 3. 実践法としてのデータ科学
 4. データ生成観測メカニズムの確率モデルからの機械学習法

1. 今更ながらデータサイエンスとは何か？

1-1. 社会変革とデータサイエンス

1-2. データサイエンスの定義

- データサイエンスの様々な定義
- データサイエンス登場の歴史

1-3. 科学的方法としてのデータサイエンス

1-4. 意思決定写像を軸としたデータサイエンスの統一視点

データサイエンス・AIと社会変革

国の動き

内閣府：

Society 5.0 第4次産業革命「IoT, AI, ビッグデータ, ロボット」

サイバー空間（仮想空間）とフィジカル空間（現実空間）の融合

AI戦略 2022

文科省：数理・データサイエンス・AI教育プログラム（認定制度）

厚生労働省：専門実践教育訓練指定講座

第4次産業革命スキル習得講座（クラウド・IoT・AI・データサイエンス、セキュリティなど）

総務省：公的統計マイクロデータの研究活用推進

大学等研究機関における各府省データのオープン利用

経済産業省：「IoT, AI, ビッグデータに関する経済産業省の取り組み」

第4次産業革命, 先進的IoT産業モデルの創出 etc.

Data science とは？

Data science

Z

"Data Scientist: The Sexiest Job of the 21st Century"
in Harvard Business Review article (2012)

The term "data science" became a buzzword.
It is now often used interchangeably with earlier
concepts like business analytics, business intelligence
etc..

英語版Wikipediaから抜粋

日本において社会一般では「データサイエンス」は
このbuzzword的な
ビジネス用語として最初は使用されはじめた

ビジネス用語
価値創造

Data science とは？（誤解への専門家からの注意）

Data science

D David Donoho (Stanford professor 2015) rejected **three simplistic and misleading definitions of data science.**

First, data science does **not** equate to **big data**, in that the size of the data set is not a criterion to distinguish data science and statistics.

Second, data science is **not** defined by the **computing skills** of sorting big data sets.

Third, data science is **a heavily applied field** where academic programs right now do not sufficiently prepare data scientists for the jobs.

英語版Wikipediaから抜粋

データサイエンスの定義の誤解

データサイエンスはビッグデータではない

データサイエンスはコンピュータースキルではない

データサイエンスは応用が重要である

Data science とは？（多様な定義が存在）

Data science

A Data science is **a multi-disciplinary field**

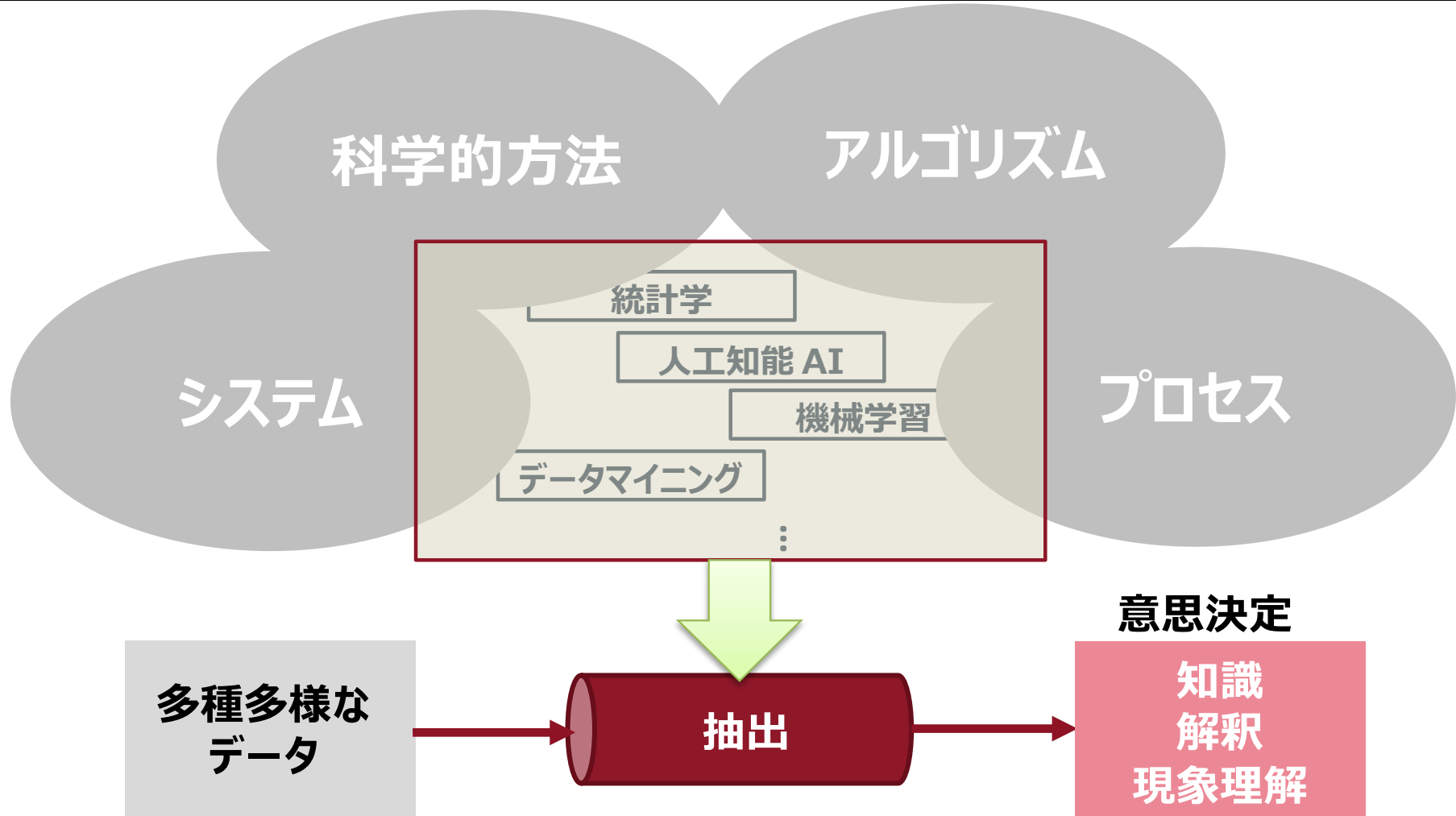
that uses scientific methods, processes, algorithms and systems to extract knowledge and insights **from data** in various forms, both structured and unstructured, similar to **data mining**

B Data science is **a concept to unify**

statistics, data analysis, machine learning and their related methods in order to understand and analyze actual phenomena **with data**.

いずれも英語版Wikipedia
から抜粋

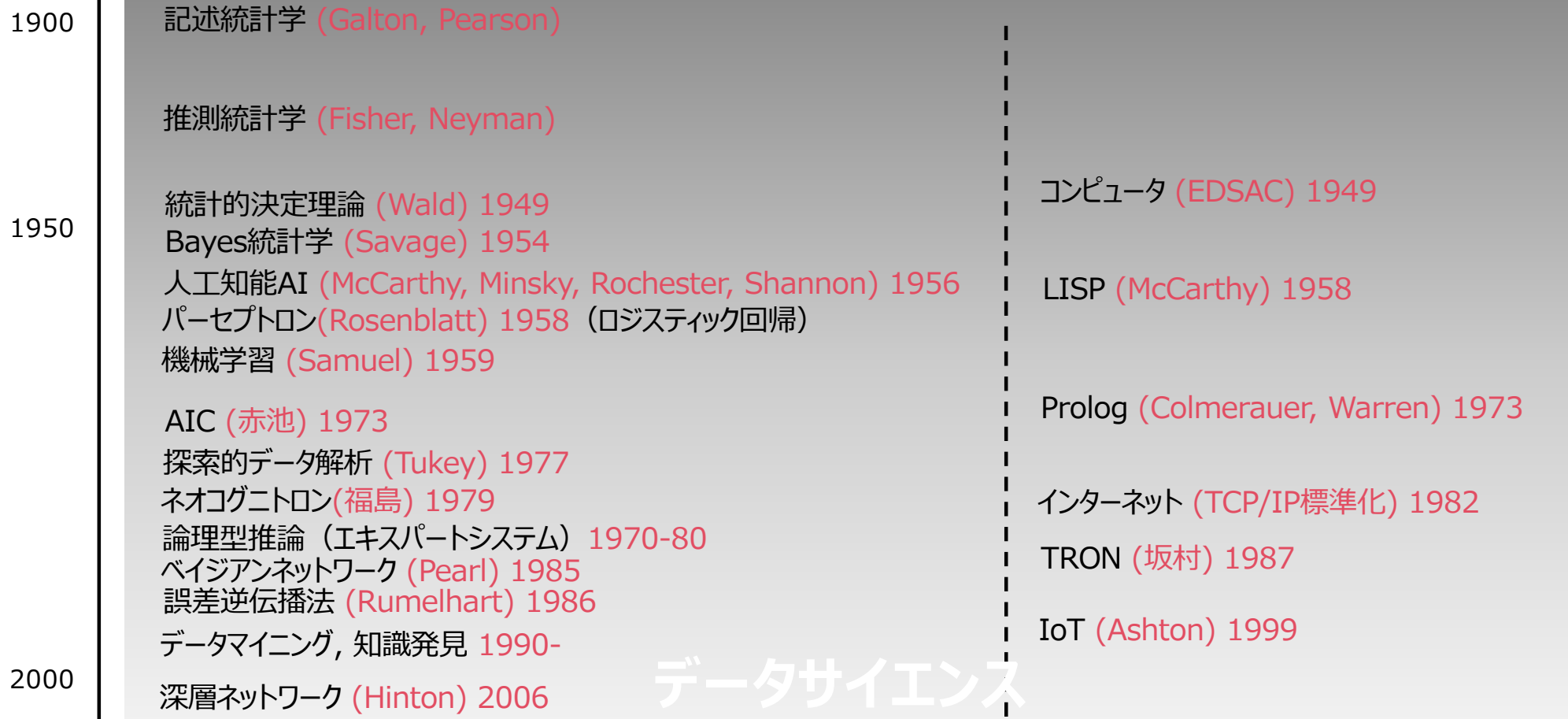
上記定義からの データサイエンスの機能と関連領域イメージ図



データサイエンス登場までの歴史

推論法 (科学的方法)

ITC技術・インフラ



データサイエンス

データサイエンス登場までの歴史 (人工知能 : AI)

コンピュータの登場

人工知能 (Artificial Intelligence)

ダートマス会議 (1956夏)

“A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON **ARTIFICIAL INTELLIGENCE**”

J. McCarthy, M. L. Minsky, N. Rochester and C.E. Shannon

人間の学習等の知的機能(知能)をコンピュータにより実現する試み

目的 : **知能とは何かをコンピュータに模倣させて探求**

あるアルゴリズムによるコンピュータの動きが人間に似ていたら, それが知能の仕組みと考える

逆にどんなアルゴリズムでも実現できた**高度な知的情報処理**は様々な応用が可能 (目的の変様)

データサイエンス登場までの歴史 (機械学習)

機械学習 (machine learning)

1959年

データを与えることでコンピュータがあるタスクを自動的に行えるようにすることを
目指す機械学習の研究をA.Samuel が提案

人間の学習機能をコンピュータで模倣する試み
知能全体の模倣を目指すAIの一分野

AIの目的の工学的なシフト ⇒ 音声や画像のパターン認識等の工学的研究へ

データサイエンス登場までの歴史 (データマイニング)

データマイニング (data mining)

知識発見 (knowledge discovery)

1990年前後

データベースの研究分野から、データから知識を発見・抽出するという意味で data mining (データマイニング), knowledge discovery (知識発見) が登場

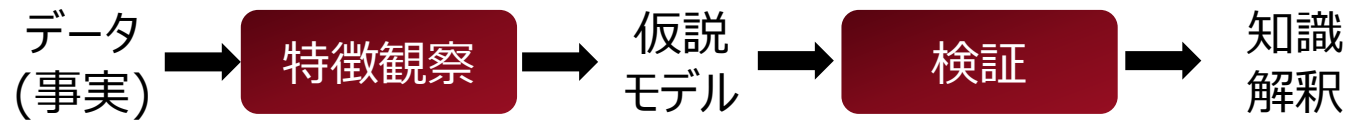
初期の頃は相関ルール抽出や簡単な決定木程度の分析法であった

帰納(事実から法則を導出する推論)を強調した用語ではあるが
学習は帰納を含むと捉えることもできる

現在は扱われる問題や使われる手法やアルゴリズムも機械学習とほぼ同じ

仮説構築検証型（科学の文法：科学的方法）

Pearson "Grammar of Science" 1892



記述統計

推測統計

仮説構築の発展（科学的方法）



1892

記述統計

Pearson

1977

探索的データ解析

Tukey

可視化

1990-

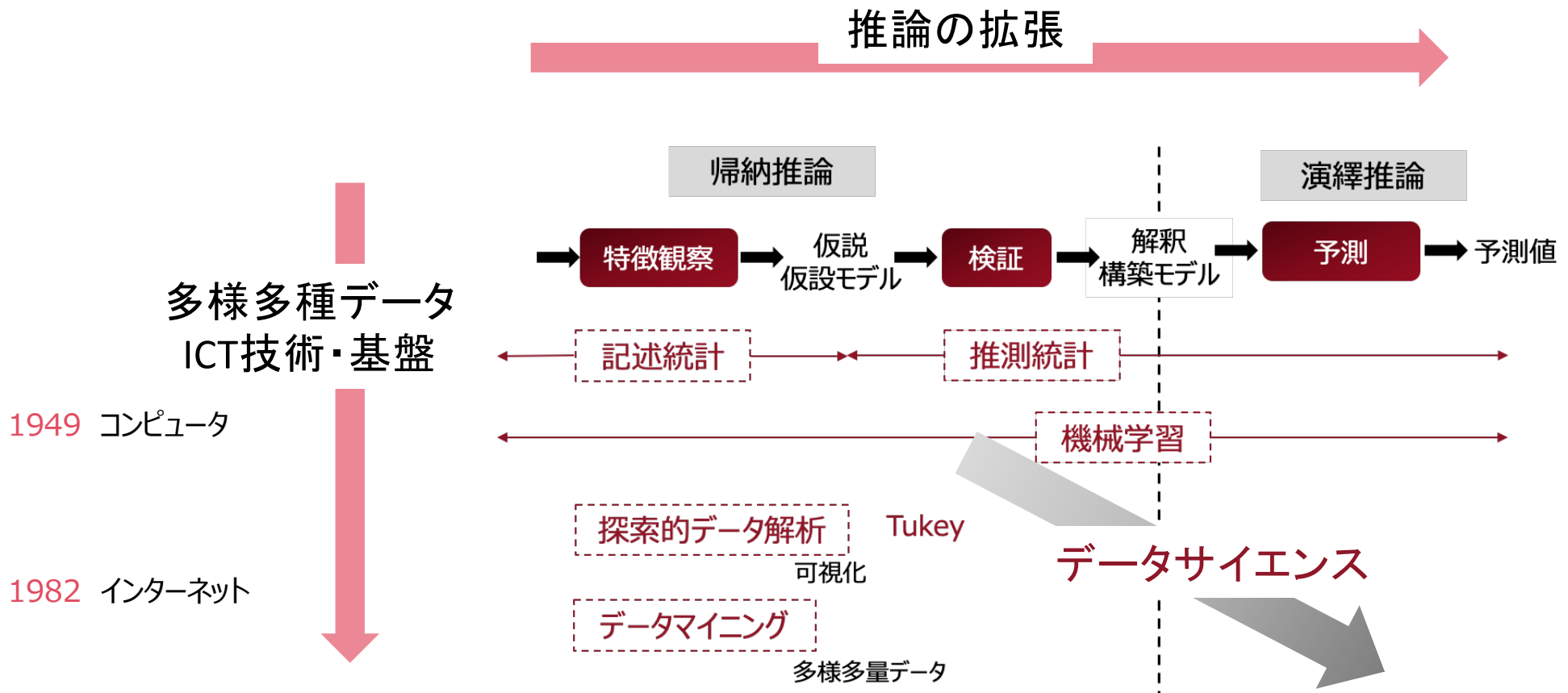
データマイニング

多様多量データ

コンピュータ (EDSAC) 1949

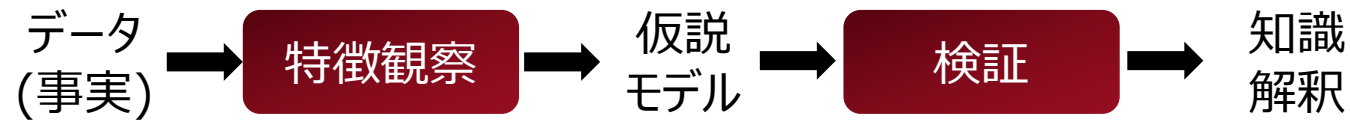
インターネット (TCP/IP標準化) 1982

2方向の拡張とデータサイエンス



仮説構築検証型（科学の文法：科学的方法）

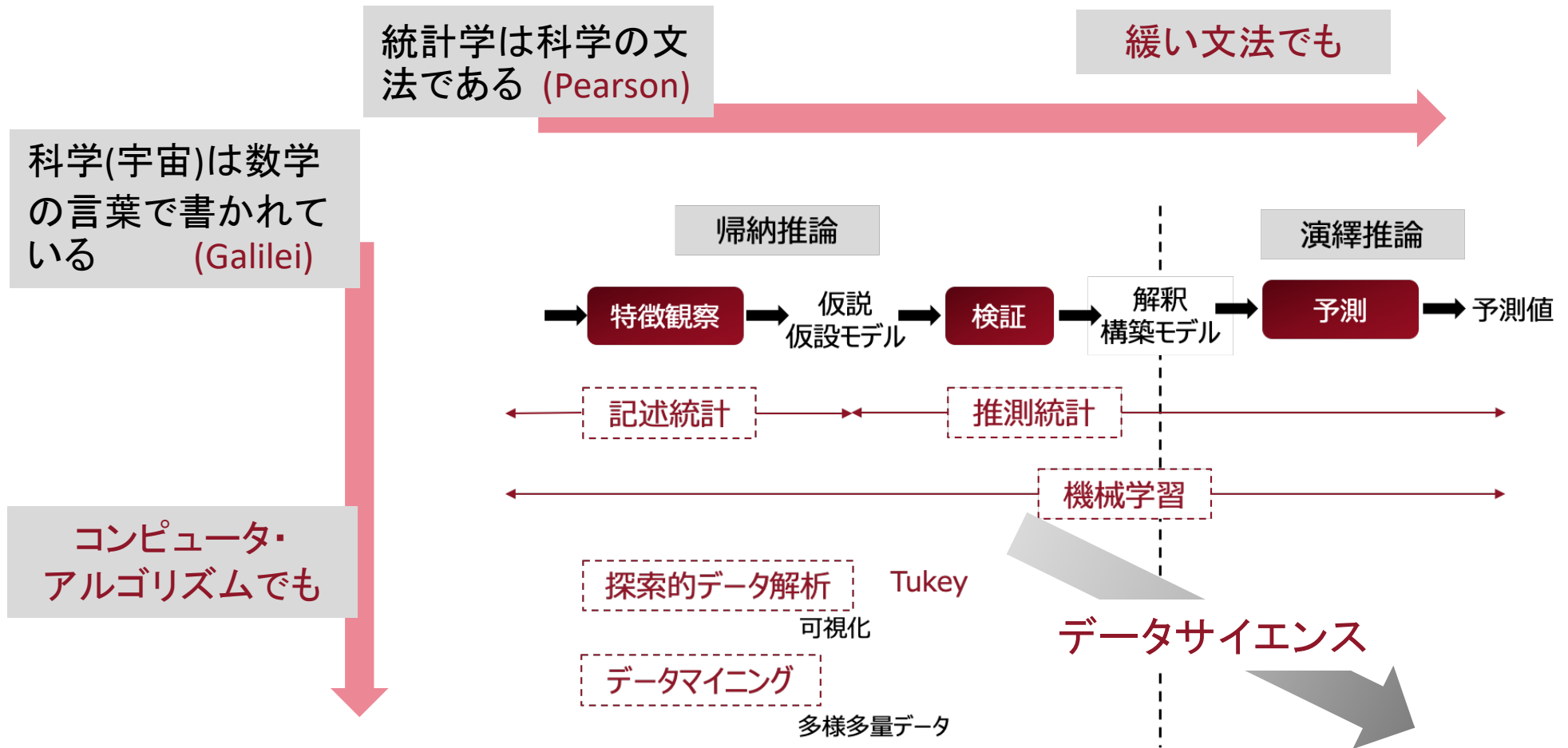
Pearson “Grammar of Science” 1892



記述統計

推測統計

2方向の拡張とデータサイエンス



科学とは？ 科学的方法とは？

科学とは？

「科学」とは、定義は様々であり哲学的な面も含んでしまう難しい問題
ここでは「**科学的方法**によって得られた知の体系」と定義

これでは議論のすり替えにみえるが、
科学的方法によって導かれた知識であることが科学を特徴づける重要な要件
多くの**科学の定義**に現れている

科学的方法とは？

「科学的方法」の一つの定義：
「**事実/証拠/データ**から**論理的な推論**により**明確な結論**を得る方法」

仮説構築検証型の**科学的方法**に限定しないより広い定義

科学とは？ 科学的方法とは？

科学とは？

「科学」とは、定義は様々であり哲学的な面も含んでしまう難しい問題
ここでは「**科学的方法**によって得られた知の体系」と定義

これでは議論のすり替えにみえるが、
科学的方法によって導かれた知識であることが科学を特徴づける重要な要件
多くの**科学の定義**に現れている

科学的方法とは？

「科学的方法」の一つの定義：
「**事実/証拠/データ**から**論理的な推論**により**明確な結論**を得る方法」

自然科学や**工学**の分野 ← 仮説検証型の科学的方法が主に適用される（20世紀に発展）
しかし、
人文科学も**社会科学**の分野も ← **データの獲得と数理的方法による論理的推論**が
可能ならば、広義の科学的方法が適用でき、それは望ましいのではないか

科学的方法の対象分野の急激な拡大（これを担うのがデータ科学）

科学的方法とは？

「科学的方法」の一つの定義：

「**事実/証拠/データから論理的推論により明確な結論を得る方法**」

データ収集：

データの種類の多様化

（数値、テキスト、音声、画像等）

データが収集できる範囲と量の拡大

（インターネット、センサー等の情報処理・通信技術の進歩とインフラの発展）

論理的推論：

多種多様多量のデータから論理的に推論する数理的理論と技術の発達

（統計学をはじめAI、機械学習、データ分析、データマイニング等の理論・技術の発展）

以前は不可能であった広範な**科学・学問領域**や様々な**意思決定問題**においても

データを用いた科学的方法が適用可能

科学的方法の対象分野の急激な拡大（これを担うのがデータ科学）



最近のノーベル賞受賞者とデータサイエンス（経済学賞）

2021年 経済学賞（因果推論による経済分析）

ガイド・インベンス（Guido Wilhelmus Imbens）

ヨシュア・アングリスト（Joshua David Angrist）

2020年 経済学賞（オークション理論）

ポール・ロバート・ミルグロム（Paul Robert Milgrom）（統計学修士）

ロバート・バトラー・ウィルソン（Robert Butler Wilson）

2019年 経済学賞（因果推論による政策評価）

アビジット・ヴィナヤック・バナジー（Abhijit Vinayak Banerjee）

エステル・デュフロ（Esther Duflo）

マイケル・ロバート・クレーマー（Michael Robert Kremer）



最近のノーベル賞受賞者とデータサイエンス（物理学賞）

2021年 物理学賞（気候シミュレーションモデル，地球温暖化）

真鍋 淑郎

クラウス・フェルディナンド・ハッセルマン（Klaus Ferdinand Hasselmann）

ジョルジョ・パリージ（Giorgio Parisi）

2020年 物理学賞（ブラックホール発見，高空間分解能イメージング分析）

ロジャー・ペンローズ（Sir Roger Penrose）

ラインハルト・ゲンツェル（Reinhard Genzel）

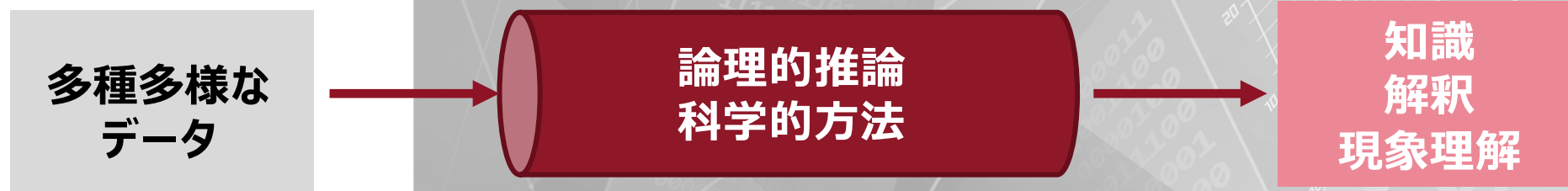
アンドレア・ミア・ゲズ（Andrea Mia Ghez）



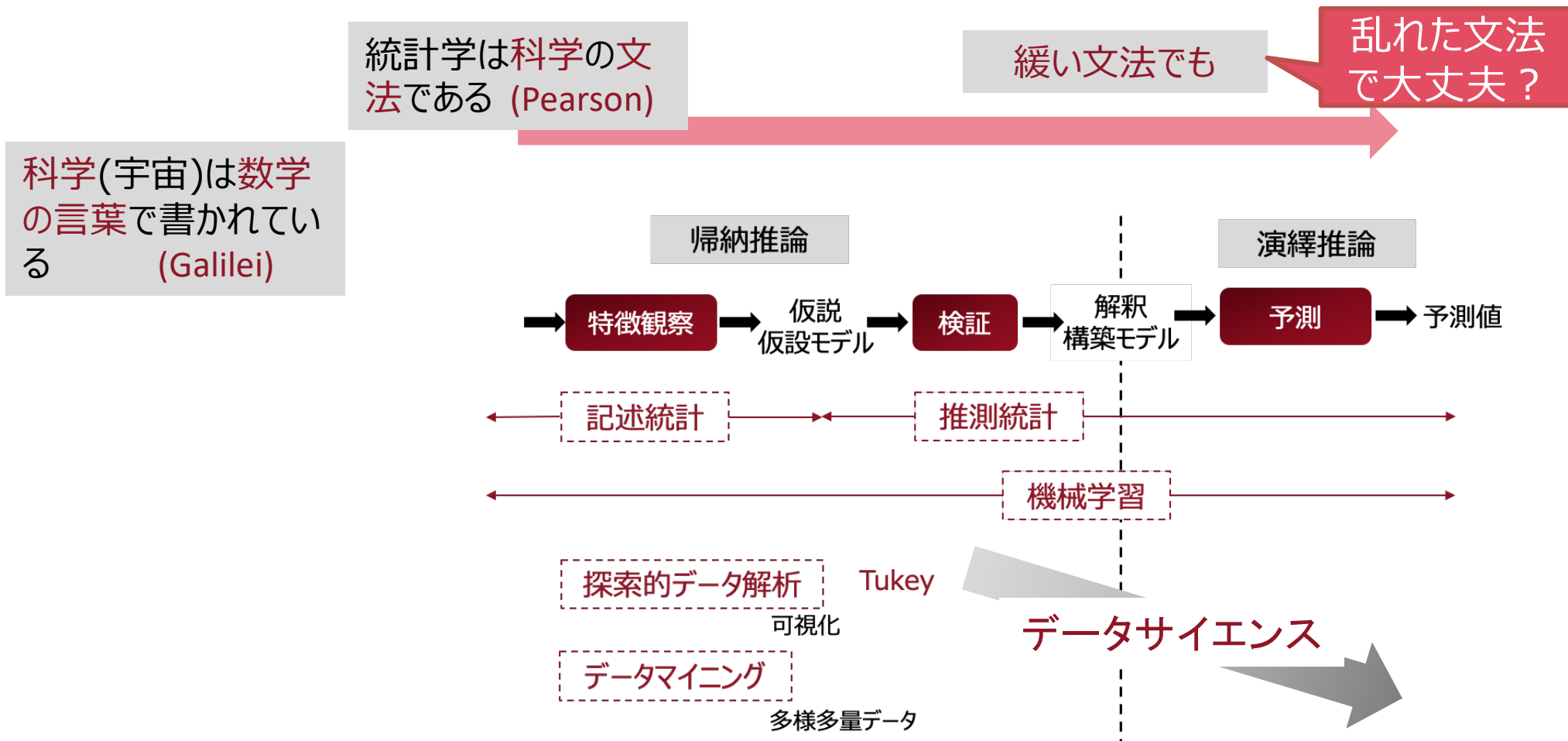
早稲田大学データ科学センターのデータサイエンス（データ科学）の定義

データ科学の定義：

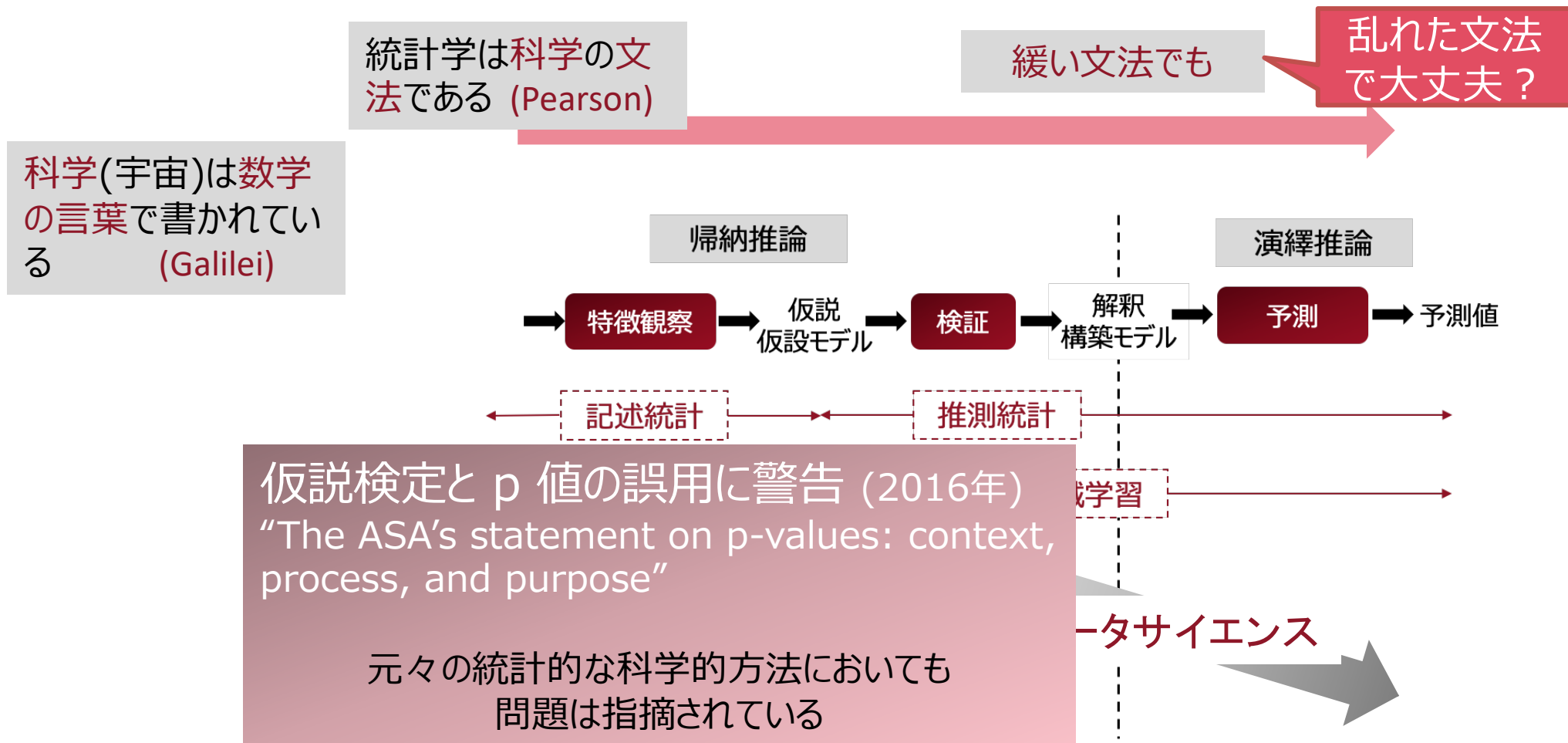
データからの論理的な意思決定の科学
or
データからの科学的方法による意思決定の科学



2方向の拡張は科学的方法として大丈夫？



2方向の拡張は科学的方法として大丈夫？



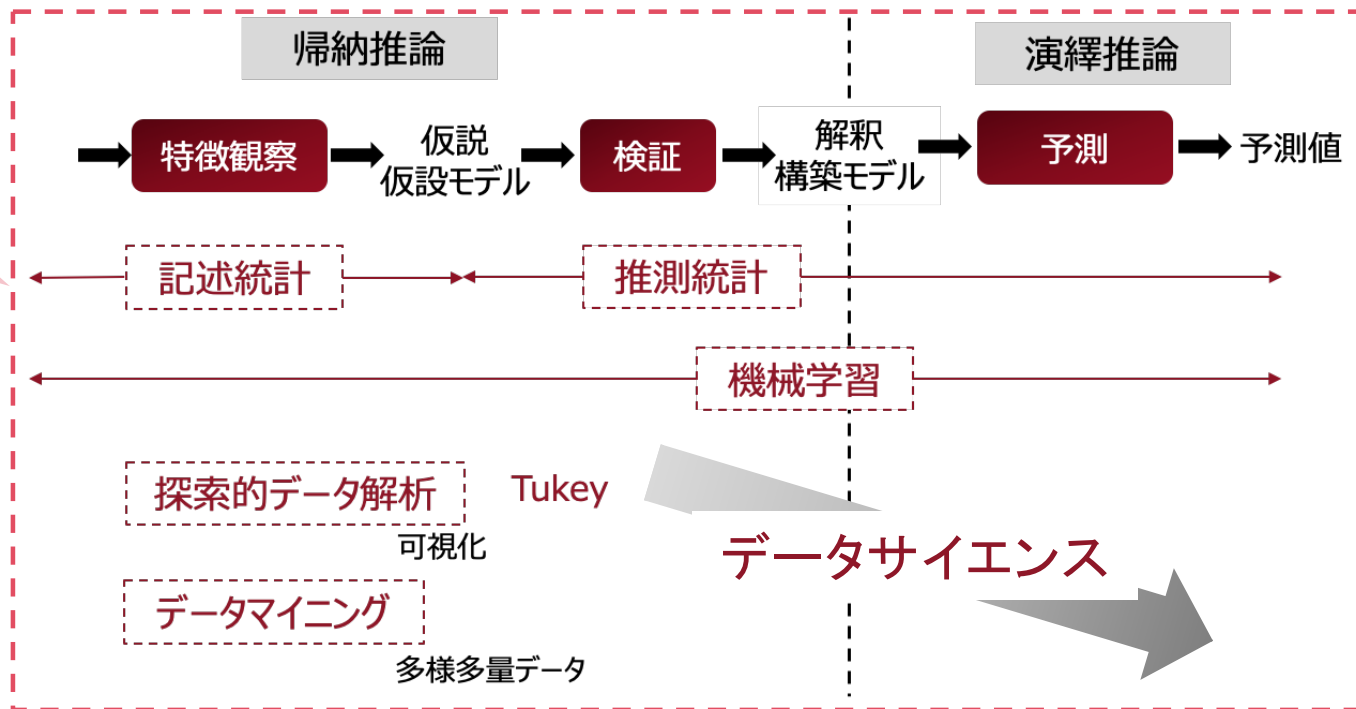
拡張された推論は論理的推論か？（統計的方法でも問題あり）

統計学は科学の文法である (Pearson)

緩い文法でも

乱れた文法で大丈夫？

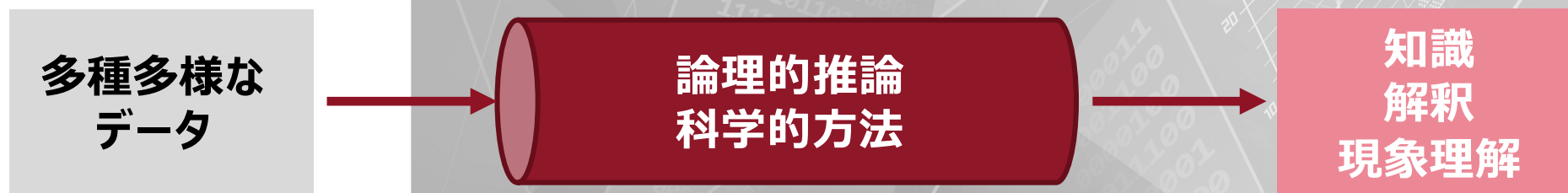
論理的推論として扱えるか (統一的に)



早稲田大データ科学センターのデータサイエンス（データ科学）の定義（再掲）

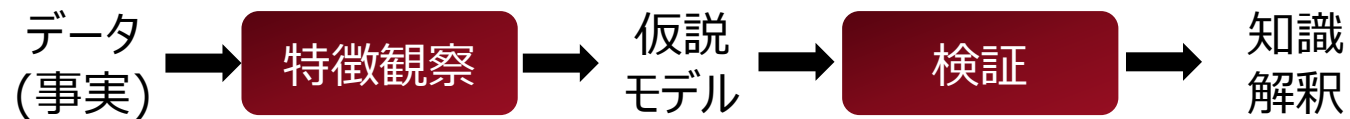
データ科学の定義：

データからの論理的な意思決定の科学
or
データからの科学的方法による意思決定の科学



仮説構築検証型（科学の文法：科学的方法）

Pearson “Grammar of Science” 1892



- 推論の方法（プロセス）が明確
- 推論の基準が明確
- 推論の一意性

（推論の健全性）

データ科学の意思決定プロセスを写像として捉える



入力

量的データ, 質的データ
ベクトルデータ, 行列データ, ...

説明するデータ, 説明されるデータ
時間的データ, 空間的データ, ...
数値データ, 言語データ, 画像データ, ...
etc.

出力

ヒストグラム
算術平均
分布パラメータ推定値
{帰無仮説, 対立仮説}
新製品を{販売を開始する, 販売を延期する}
新たな目的変数の予測値
選択された統計モデル
分割領域, 決定木
etc.

意思決定写像を軸とした論理性の解釈（統一的視点に向けて）

写像が一意

||

データから一意な意思決定
(客観的, 手続きが明確)

明確な設定・基準のものと
論理的に写像が
導かれている

写像を一意にするための制約を明示

目的:
設定:
評価基準:

データ集合
 $x \in \mathcal{A}$

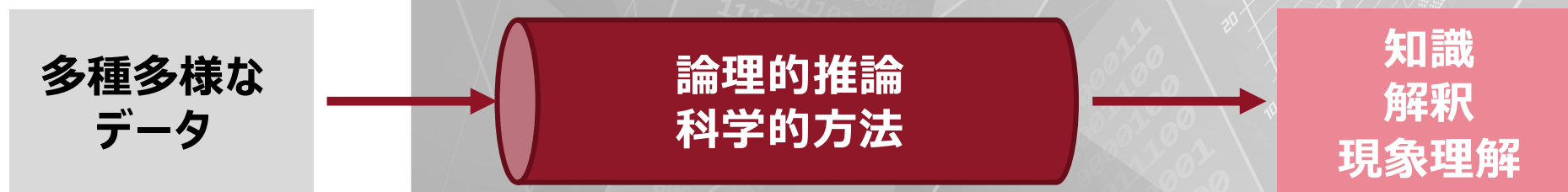
意思決定写像

決定集合
 $d \in \mathcal{B}$

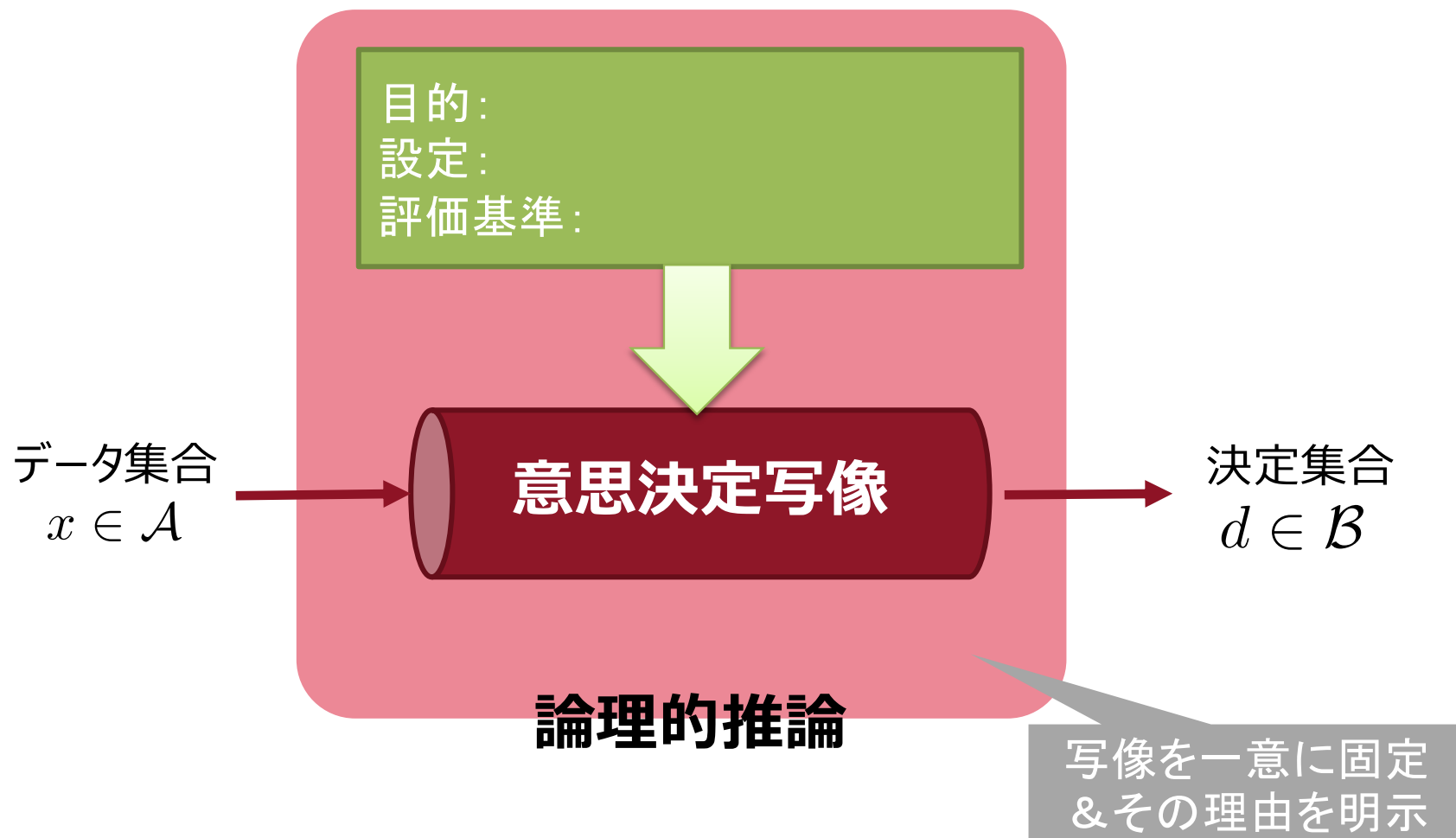
早稲田大データ科学センターのデータサイエンス（データ科学）の定義（再掲）

データ科学の定義：

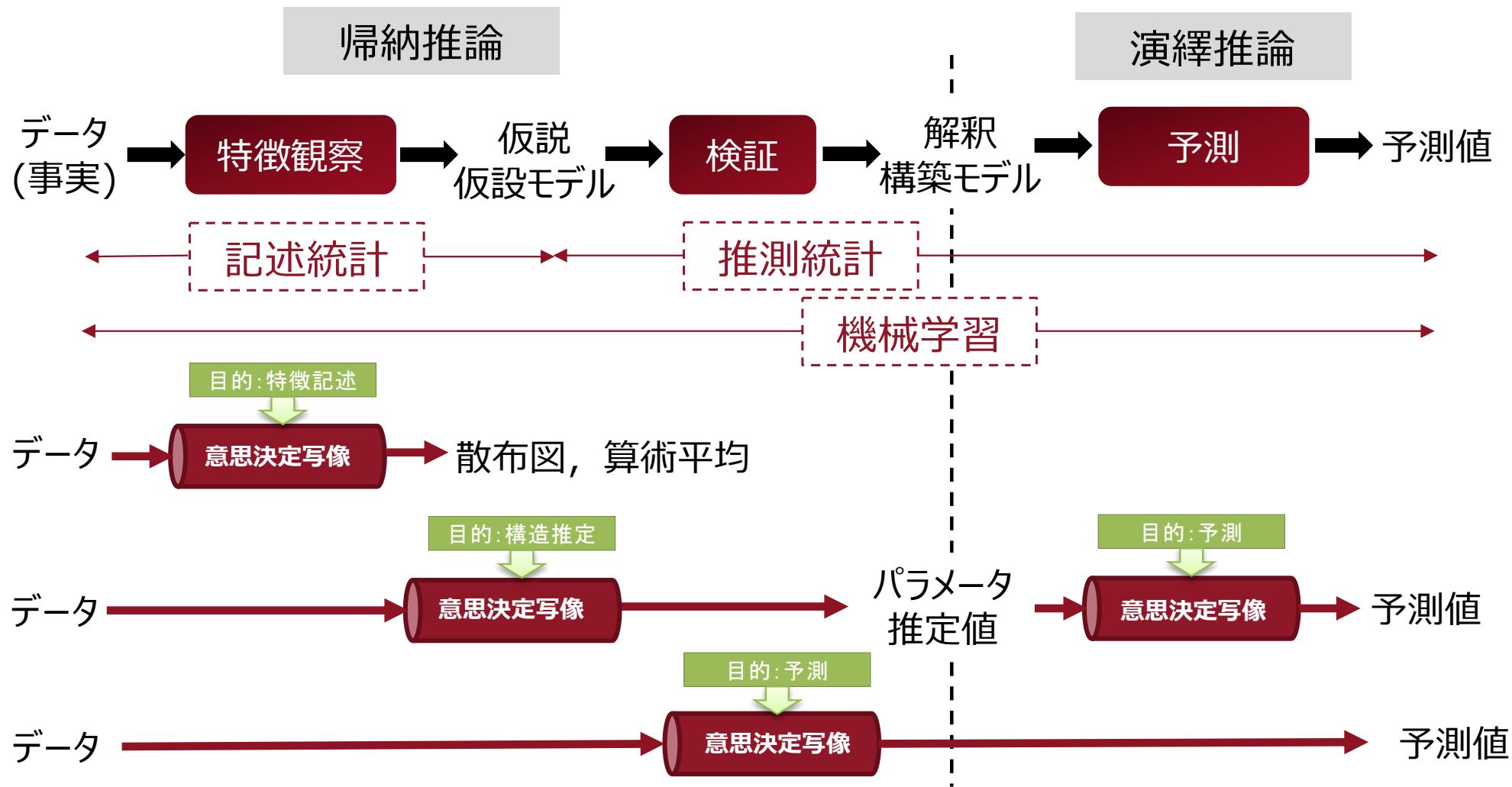
データからの論理的な意思決定の科学
or
データからの科学的方法による意思決定の科学



意思決定写像を軸とした統一の視点



様々な推論を意思決定写像で捉える（目的の明確化）



意思決定写像の例（構造推定としての回帰問題の例）

回帰係数も確率変数と設定した場合の意思決定写像の例

目的: 説明変数 x_1, \dots, x_n が与えられたもとでの y_1, \dots, y_n の確率的データ生成観測メカニズムを明らかにしたい

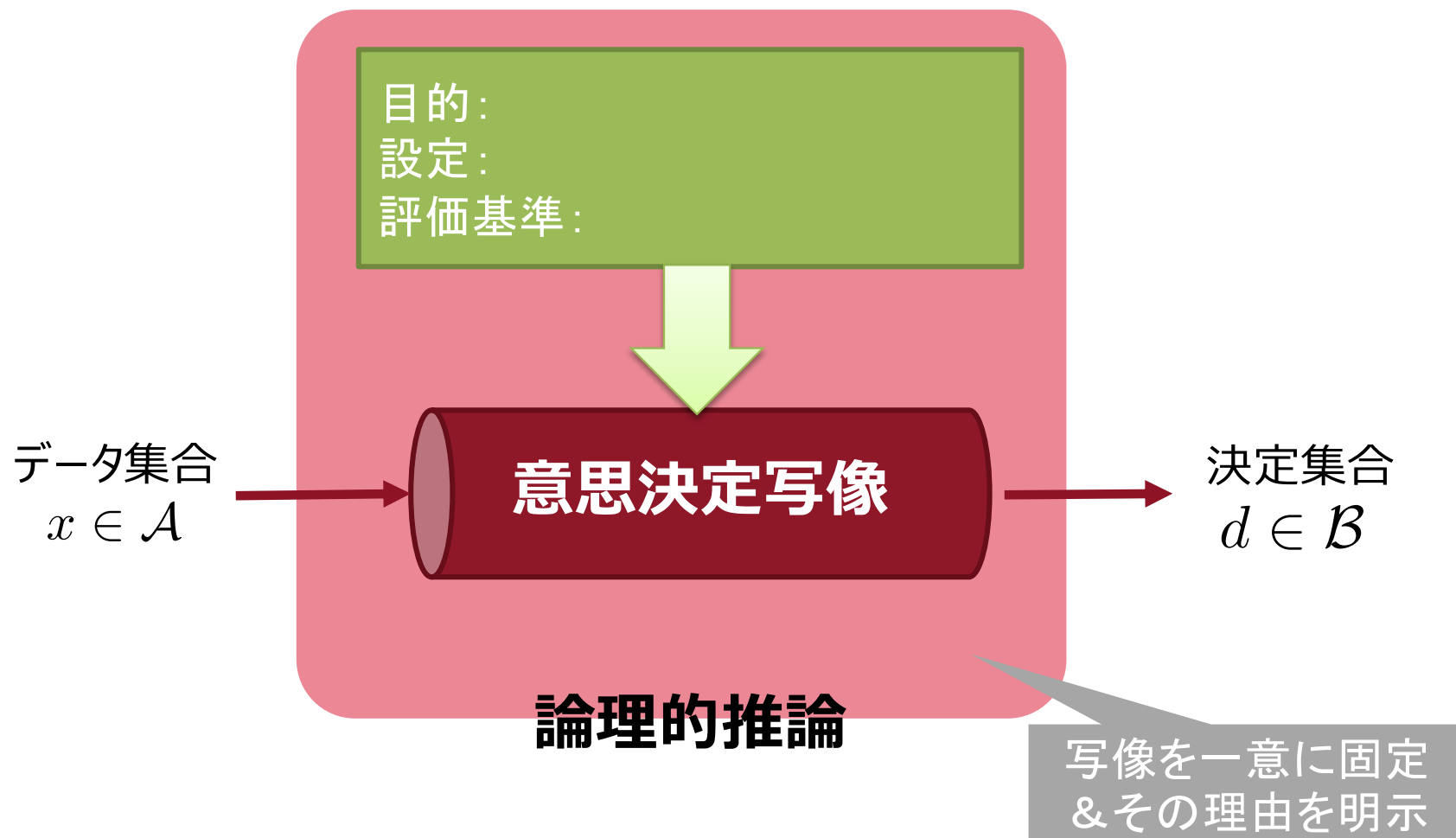
設定: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$
 ε_i は i.i.d. で正規分布 $\mathcal{N}(0, \sigma_\varepsilon^2)$ に従う
 β_0, β_1 は事前分布 $p(\beta_0, \beta_1)$ に従う

損失関数: $\ell(\beta_0, \beta_1, d((x_1, y_1), \dots, (x_n, y_n)))$

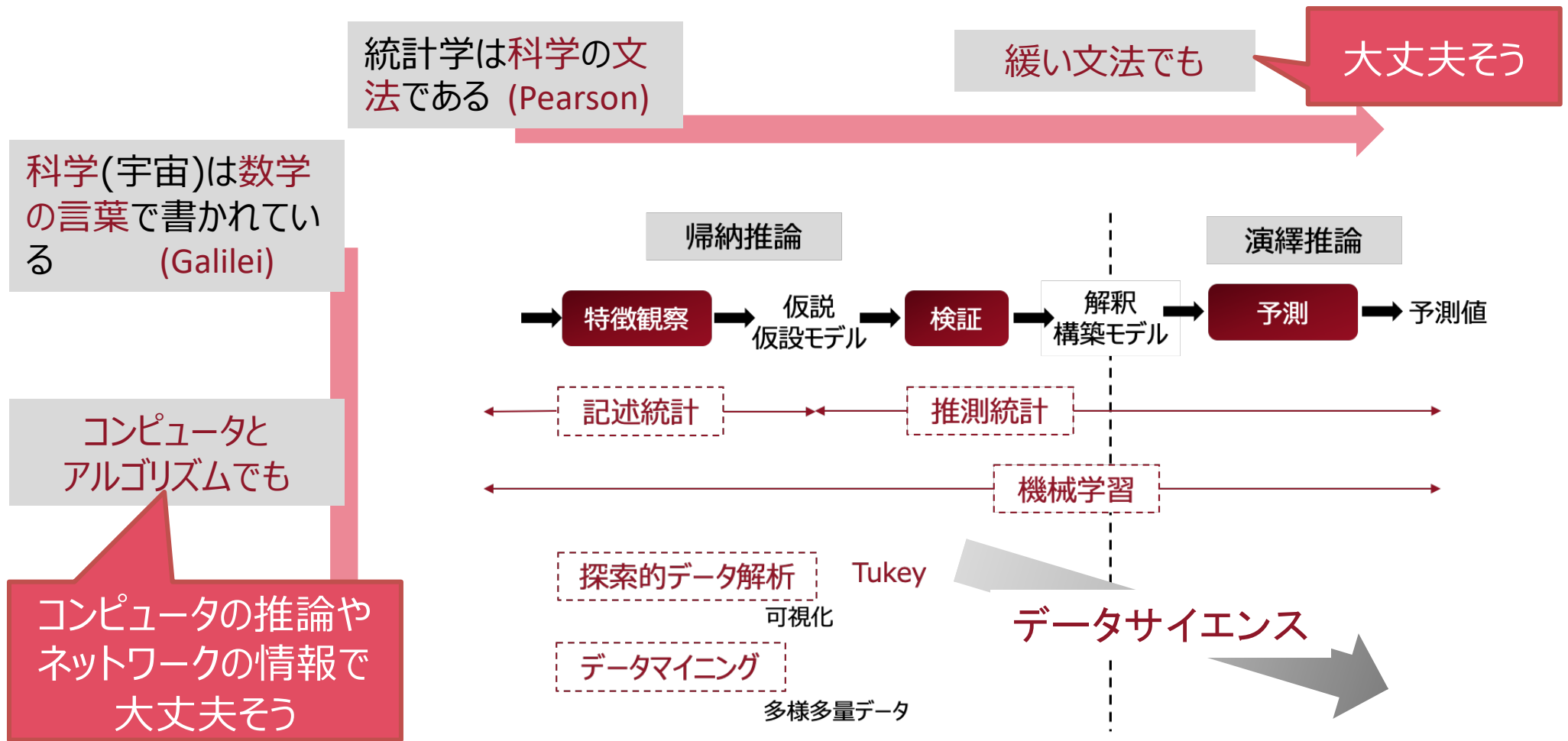
評価基準: ベイズ危険関数最小化



意思決定写像を軸とした統一の視点



2方向の拡張とデータサイエンス



早稲田大学データ科学センターのご案内



CDS 早稲田大学
データ科学センター

お問い合わせ先

【ホームページ】

<https://www.waseda.jp/inst/cds/>



【電話】

03-3204-9164

【E-mail】

cds-info@list.waseda.jp

【担当】

梅原事務長、田尻、草野、松木、
高橋、箱川

交通アクセス

169-8050 東京都新宿区西早稲田1-6-1 1号館3階
東京メトロ東西線 早稲田駅 徒歩5分

