

ポスト選択推論による データ駆動仮説の統計的信頼性評価

竹内 一郎

名古屋工業大学・情報工学専攻・機械学習 & データ科学研究室

理化学研究所・革新知能統合研究センター・データ駆動型生物医科学チーム

自己紹介

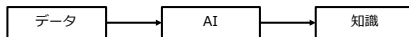
- ▶ 所属：名古屋工業大学・情報工学専攻「機械学習＆データ科学研究室」
- ▶ 兼任：理化学研究所・革新知能統合研究センター「データ駆動型生物医科学チーム」
- ▶ 専門：機械学習
- ▶ 研究：データ駆動型科学の理論と実践
- ▶ プロジェクト：データ駆動型の生物科学，医療科学，材料科学

AI 活用におけるアクセルとブレーキ

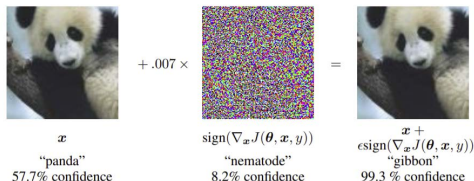
- ▶ AI の ELSI (Ethical, Legal and Social Issues)
 - ▶ 公平性
 - ▶ 適用先 (例: 軍事, SDG)
- ▶ AI の解釈性
 - ▶ 可視化
 - ▶ ルール抽出
- ▶ AI の信頼性
 - ▶ ロバスト性
 - ▶ 統計的有意性

AI による発見の信頼性

- ▶ AI によって発見された知識は信頼できるのか？

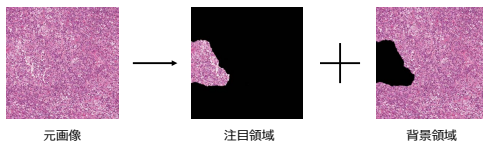


- ▶ ロバスト性 (robustness)



Goodfellow et al. (ICLR2015) Fig.1

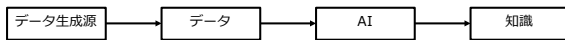
- ▶ 統計的有意性 (statistical significance)



JST CREST 信頼される AI システム領域「AI 駆動仮説の静的・動的信頼性保証と医療への展開」

データ発生源の確率モデル

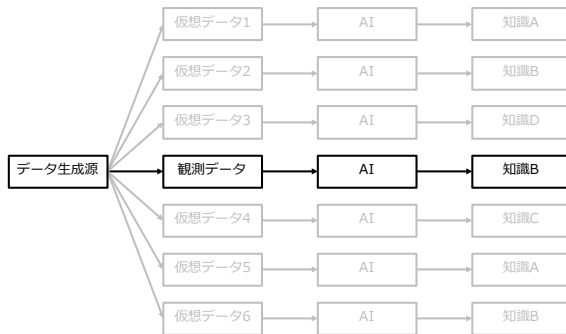
▶ データ発生源の確率モデル



- ▶ ロバスト性（最悪の場合はどうなるか）
- ▶ 統計的有意性（偶然の産物ではないのか）

データ発生源の確率モデル

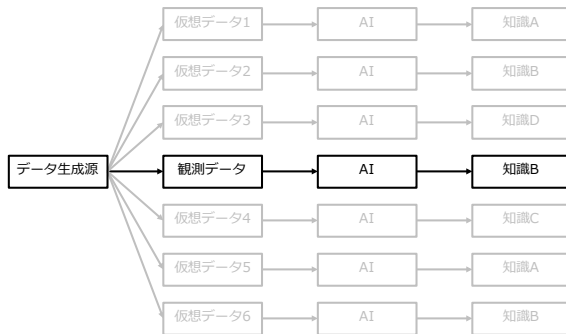
▶ データ発生源の確率モデル



- ▶ ロバスト性（最悪の場合はどうなるか）
- ▶ 統計的有意性（偶然の産物ではないのか）

データ発生源の確率モデル

▶ データ発生源の確率モデル

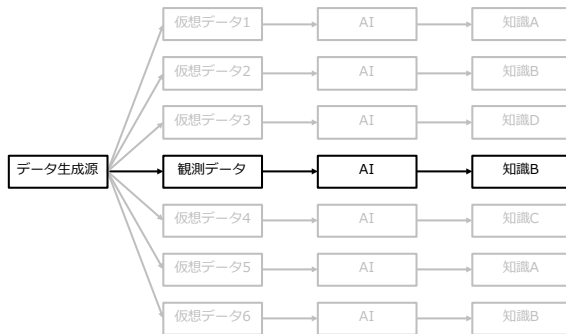


▶ ロバスト性（最悪の場合はどうなるか）

▶ 統計的有意性（偶然の産物ではないのか）

データ発生源の確率モデル

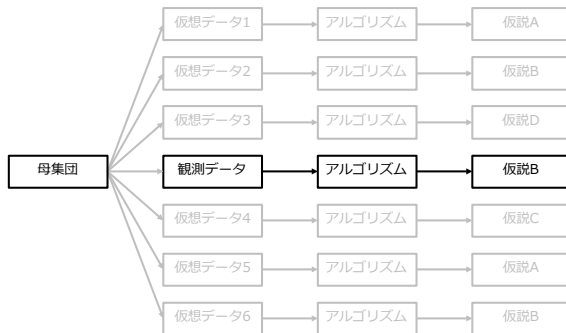
▶ データ発生源の確率モデル



- ▶ ロバスト性（最悪の場合はどうなるか）
- ▶ 統計的有意性（偶然の産物ではないのか）

データ発生源の確率モデル

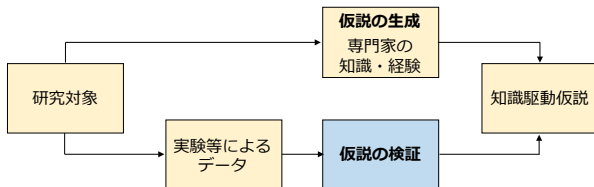
▶ データ発生源の確率モデル



- ▶ ロバスト性（最悪の場合はどうなるか）
- ▶ 統計的有意性（偶然の産物ではないのか）

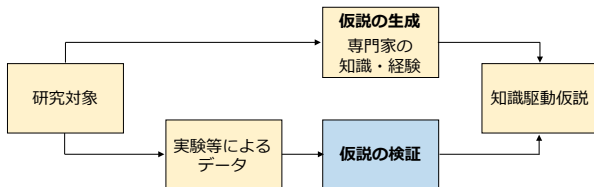
知識駆動仮説とデータ駆動仮説

(伝統的な) 科学研究における仮説検証の枠組

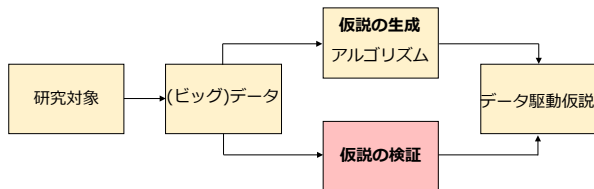


知識駆動仮説とデータ駆動仮説

(伝統的な) 科学研究における仮説検証の枠組



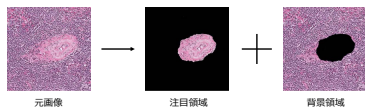
データ駆動科学における仮説検証の枠組



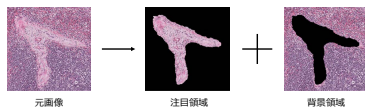
本講演で紹介する研究成果の例

ポスト選択推論：「AI により発見された仮説の統計的有意性を正しく定量化する方法」

▶ 真に異常領域が存在する場合

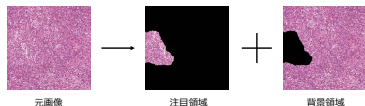


通常の p 値： $\underbrace{0.000}_{\text{T.P.}}$ 新たな p 値： $\underbrace{0.000}_{\text{T.P.}}$

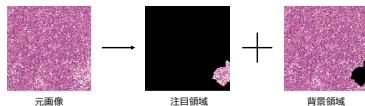


通常の p 値： $\underbrace{0.000}_{\text{T.P.}}$ 新たな p 値： $\underbrace{0.000}_{\text{T.P.}}$

▶ 真に異常領域が存在しない場合



通常の p 値： $\underbrace{0.000}_{\text{F.P.}}$ 新たな p 値： $\underbrace{0.351}_{\text{T.N.}}$



通常の p 値： $\underbrace{0.000}_{\text{F.P.}}$ 新たな p 値： $\underbrace{0.734}_{\text{T.N.}}$

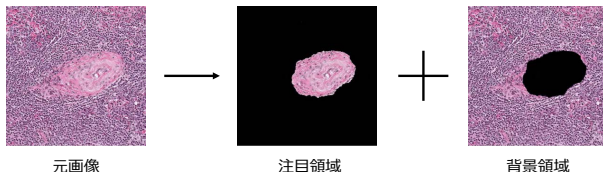
本講演のアウトライン

- ▶ Part 1: 仮説選択バイアスと選択的推論
- ▶ Part 2: ポスト選択推論（条件付選択的推論）
- ▶ Part 3: 深層学習モデルのポスト選択推論

Part 1: 仮説選択バイアスと選択的推論

例題 1：医療画像セグメンテーション

- ▶ 病理画像からセグメンテーションにより注目領域を発見したい



- ▶ 画素値 (n 個) を並べたベクトル $X \in \mathbb{R}^n$

$$\underbrace{X}_{\text{画素値ベクトル}} = \underbrace{M}_{\text{真の画素値ベクトル}} + \underbrace{\varepsilon}_{\text{ノイズベクトル}}, \quad \underbrace{\varepsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)}_{\text{ノイズは正規分布に従う}}$$

- ▶ セグメンテーションアルゴリズム $\mathcal{A}$

$$\underbrace{\mathcal{A}}_{\text{アルゴリズム}} : \underbrace{X}_{\text{画素値ベクトル}} \mapsto \left\{ \underbrace{\mathcal{O}_X}_{\text{注目領域の画素集合}}, \underbrace{B_X}_{\text{背景領域の画素集合}} \right\}$$

統計的仮説検証

▶ 統計的仮説検証

▶ 帰無仮説

$$H_0 : \underbrace{\frac{1}{|\mathcal{O}_X|} \sum_{i \in \mathcal{O}_X} M_i}_{\text{注目領域の画素値の平均}} = \underbrace{\frac{1}{|\mathcal{B}_X|} \sum_{i \in \mathcal{B}_X} M_i}_{\text{背景領域の画素値の平均}}$$

▶ 対立仮説

$$H_1 : \underbrace{\frac{1}{|\mathcal{O}_X|} \sum_{i \in \mathcal{O}_X} M_i}_{\text{注目領域の画素値の平均}} \neq \underbrace{\frac{1}{|\mathcal{B}_X|} \sum_{i \in \mathcal{B}_X} M_i}_{\text{背景領域の画素値の平均}}$$

▶ 検定統計量：注目領域と背景領域の画素値の平均の差

$$\Delta_X := \frac{1}{|\mathcal{O}_X|} \sum_{i \in \mathcal{O}_X} X_i - \frac{1}{|\mathcal{B}_X|} \sum_{i \in \mathcal{B}_X} X_i$$

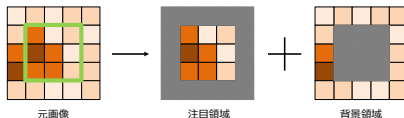
▶ 統計的有意性（両側検定の p 値）

$$p = \Pr \left(\underbrace{|\Delta_X|}_{\text{確率変数}} \geq \underbrace{|\Delta_x|}_{\text{観測値}} \right)$$

知識駆動仮説とデータ駆動仮説

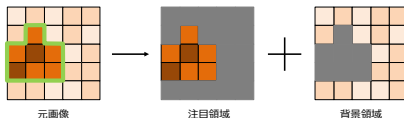
- 知識駆動仮説：注目・背景領域がデータによらない $\Rightarrow$ z 検定 or t 検定

$$\underbrace{\frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} M_i}_{\text{固定された注目領域の画素値の平均}} = \text{or } \neq \underbrace{\frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} M_i}_{\text{固定された背景領域の画素値の平均}}$$

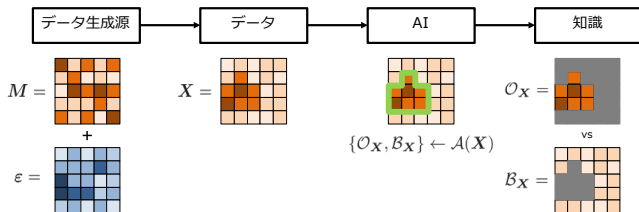


- データ駆動仮説：注目・背景領域がデータから決まる $\Rightarrow$ アルゴリズム依存

$$\underbrace{\frac{1}{|\mathcal{O}_X|} \sum_{i \in \mathcal{O}_X} M_i}_{\text{選択された注目領域の画素値の平均}} = \text{or } \neq \underbrace{\frac{1}{|\mathcal{B}_X|} \sum_{i \in \mathcal{B}_X} M_i}_{\text{選択された背景領域の画素値の平均}}$$

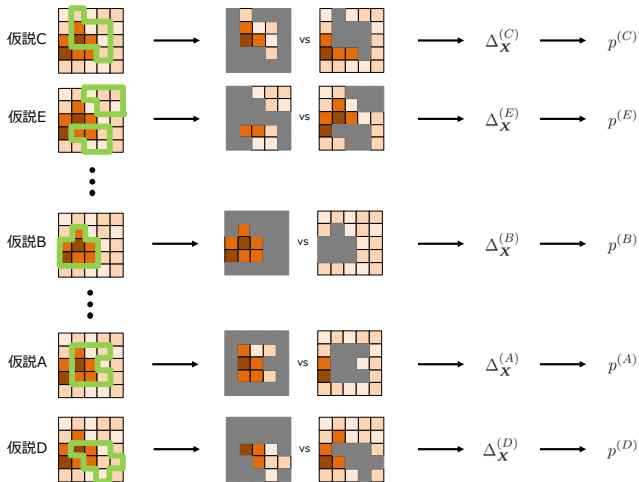


▶ AIにより発見される知識はデータ駆動



仮説の選択と選択バイアス

- ▶ 仮説の選択は多重比較問題と解釈可能

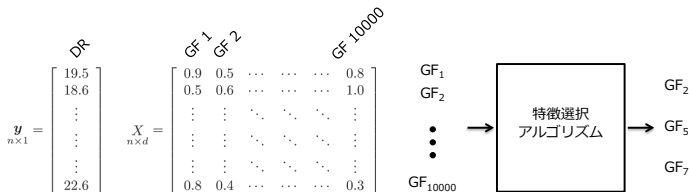


- ▶ 多くの仮説を比較して最も有意な (p 値の小さな) 仮説を選ぶときは補正が必要

例題 2：線形モデルの特徴選択

(例) 10000 個の遺伝子のうち、薬剤効果予測に有益なものを選択したい

▶ 高次元データと特徴選択



▶ 選択された特徴を用いた最小二乗方による線形モデル

$$\hat{y}_i = \hat{\beta}_2 x_{i2} + \hat{\beta}_5 x_{i5} + \hat{\beta}_7 x_{i7}$$

ただし、

$$\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_5 \\ \hat{\beta}_7 \end{bmatrix} = \underset{\beta_2, \beta_5, \beta_7}{\operatorname{argmin}} \sum_{i=1}^n (y_i - (\beta_2 x_{i2} + \beta_5 x_{i5} + \beta_7 x_{i7}))^2$$

例題 2：線形モデルの特徴選択（定式化）

- ▶ データ（例の場合は $n = 50, d = 10000$ ）

$$\mathbf{X} \in \mathbb{R}^{n \times d}, \mathbf{Y} \in \mathbb{R}^n$$

- ▶ 確率モデル

$$\underbrace{\mathbf{Y}}_{\text{薬剤効果}} = \underbrace{\mu(\mathbf{X})}_{\text{真の薬剤効果}} + \underbrace{\boldsymbol{\varepsilon}}_{\text{ノイズ}}, \quad \underbrace{\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma)}_{\text{ノイズは正規分布}}$$

- ▶ 特徴選択アルゴリズム $\mathcal{A}$

$$\mathcal{A} : \mathbf{Y} \mapsto \mathcal{M}_{\mathbf{Y}},$$

ただし、 $\mathcal{M}_{\mathbf{Y}}$ は選択された特徴の集合で、例の場合は $\mathcal{M}_{\mathbf{y}} = \{2, 5, 7\}$

- ▶ 選択された特徴のみを用いた最小二乗法による線形モデル

$$\hat{\boldsymbol{\beta}}_{\mathcal{M}_{\mathbf{Y}}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{|\mathcal{M}_{\mathbf{Y}}|}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}_{\mathcal{M}_{\mathbf{Y}}}^{\top} \boldsymbol{\beta}\|^2 = (\mathbf{X}_{\mathcal{M}_{\mathbf{Y}}}^{\top} \mathbf{X}_{\mathcal{M}_{\mathbf{Y}}})^{-1} \mathbf{X}_{\mathcal{M}_{\mathbf{Y}}}^{\top} \mathbf{Y}$$

統計的仮説検証

▶ 統計的仮説検証

▶ 母最小二乗解 (population least square solution)

$$\beta_{\mathcal{M}_Y} = (X_{\mathcal{M}_Y}^\top X_{\mathcal{M}_Y})^{-1} X_{\mathcal{M}_Y}^\top \underbrace{\mu(X)}_{\text{真の薬剤効果}}$$

▶ 帰無仮説

$$H_0 : \underbrace{\beta_{\mathcal{M}_Y, j}}_{\text{選択されたモデルにおける特徴 } j \text{ の影響}} = 0$$

▶ 対立仮説

$$H_1 : \underbrace{\beta_{\mathcal{M}_Y, j}}_{\text{選択されたモデルにおける特徴 } j \text{ の影響}} \neq 0$$

▶ 検定統計量

$$\hat{\beta}_{\mathcal{M}_Y, j} = (X_{\mathcal{M}_Y}^\top X_{\mathcal{M}_Y})^{-1} X_{\mathcal{M}_Y}^\top \underbrace{Y}_{\text{観測された薬剤効果}}$$

▶ 統計的有意性 (両側検定の p 値)

$$p = \Pr \left(\underbrace{|\hat{\beta}_{\mathcal{M}_Y, j}|}_{\text{確率変数}} \geq \underbrace{|\hat{\beta}_{\mathcal{M}_Y, j}|}_{\text{観測値}} \right)$$

知識駆動仮説とデータ駆動仮説

- ▶ 知識駆動仮説：特徴集合がデータによらない $\Rightarrow$ z 検定 or t 検定

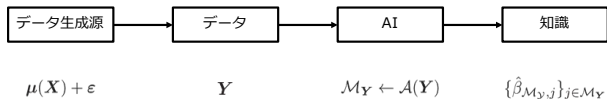
$$\underbrace{\beta_{\mathcal{M},j}}_{\text{固定されたモデルにおける特徴 } j \text{ の影響}} = \text{or } \neq 0$$

- ▶ データ駆動仮説：特徴集合がデータから決まる $\Rightarrow$ アルゴリズム依存

$$\underbrace{\beta_{\mathcal{M}_{\mathbf{Y}},j}}_{\text{選択されたモデルにおける特徴 } j \text{ の影響}} = \text{or } \neq 0$$

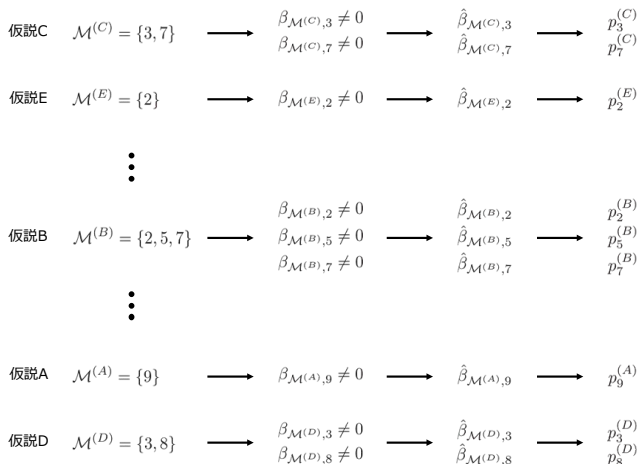
AI とデータ駆動仮説

- ▶ AI により発見される知識はデータ駆動



仮説の選択と選択バイアス

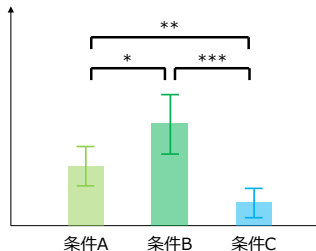
- ▶ 仮説の選択は多重比較問題と解釈可能



- ▶ 多くの仮説を比較して最も有意な (p 値の小さな) 仮説を選ぶときは補正が必要

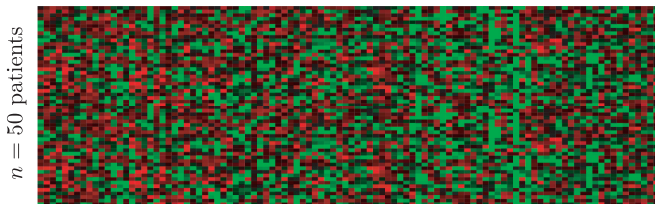
多重比較・多重検定（選択的推論）

- ▶ 伝統的な多重比較問題：数個程度の仮説候補からの選択



- ▶ 生命情報解析（2000 年～）：数千～数万個の仮説からの選択

$d = 10000$ genes



選択的推論の3つのアプローチ

- ▶ Family-wise error rate (FWER)
1 つでも誤発見をする確率を 0.05 未満に制御
- ▶ False discover rate (FDR)
発見された仮説のなかで誤発見の割合を 0.05 未満に制御
- ▶ Conditional inference (Post-selection inference)
仮説が選択された条件のもとで誤発見である確率を 0.05 未満に制御

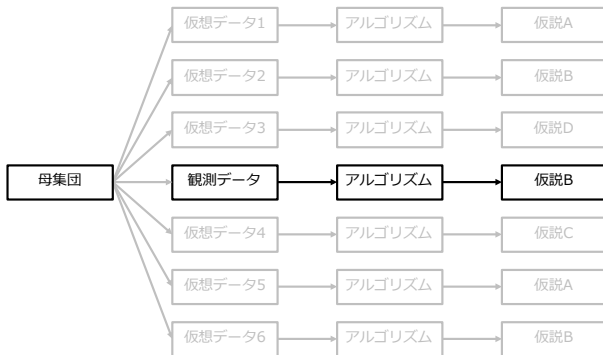
Part 1 のまとめ

- ▶ AI によって発見された仮説（知識）はデータ駆動仮説
- ▶ データ駆動仮説の統計的信頼性は伝統的な統計学では対処できない
- ▶ 多数の仮説から最適なものを選択する問題（多重比較）と解釈可能

Part 2: 条件付選択の推論（ポスト選択推論）

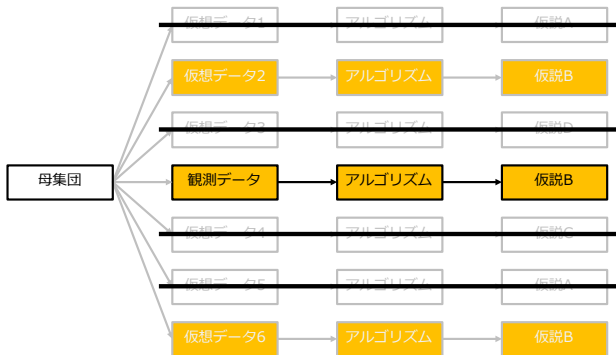
条件付き推論の概念（その1）

- ▶ 条件付推論：AI により仮説の選択された場合のランダムネスのみを考える



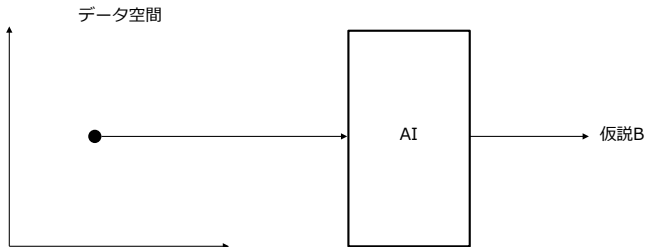
条件付き推論の概念（その1）

- ▶ 条件付推論：AI により仮説の選択された場合のランダムネスのみを考える



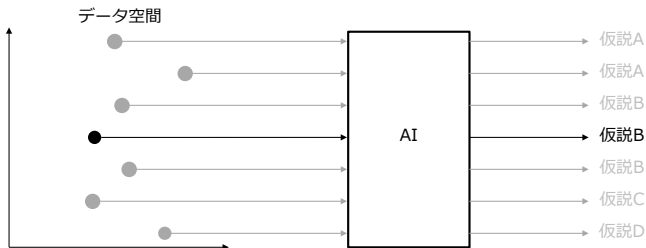
条件付き推論の概念（その2）

- ▶ 条件付推論：AI により仮説の選択された場合のランダムネスのみを考える



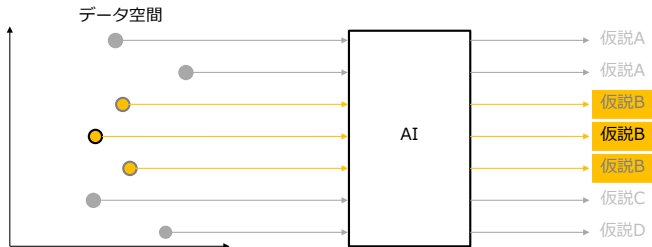
条件付き推論の概念（その2）

- ▶ 条件付推論：AI により仮説の選択された場合のランダムネスのみを考える



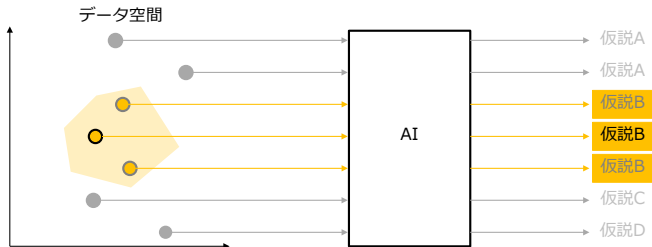
条件付き推論の概念（その2）

- ▶ 条件付推論：AI により仮説の選択された場合のランダムネスのみを考える



条件付き推論の概念（その2）

- ▶ 条件付推論：AI により仮説の選択された場合のランダムネスのみを考える



医療画像セグメンテーションにおける条件付推論

- ▶ (通常の) 統計的有意性指標

$$p = \Pr \left(\underbrace{|\Delta_X|}_{\text{確率変数}} \geq \underbrace{|\Delta_x|}_{\text{観測値}} \right)$$

- ▶ 条件付推論の統計的有意性指標

$$p = \Pr \left(\underbrace{|\Delta_X|}_{\text{確率変数}} \geq \underbrace{|\Delta_x|}_{\text{観測値}} \mid \underbrace{\{\mathcal{O}_X, \mathcal{M}_X\} = \{\mathcal{O}_x, \mathcal{B}_x\}}_{\text{注目領域と背景領域が観測値のものと同じ}} \right)$$

医療画像セグメンテーションにおける条件付推論

- ▶ 閾値 θ に基づくセグメンテーションアルゴリズム

$$\mathcal{O}_X = \{X_i \geq \theta\},$$

$$\mathcal{B}_X = \{X_i < \theta\}$$

- ▶ 仮説選択イベント

$$X_i \geq \theta, i \in \mathcal{O}_x,$$

$$X_i < \theta, i \in \mathcal{B}_x$$

- ▶ 条件付推論の統計的有意性指標

$$\begin{aligned} p &= \Pr \left(\underbrace{|\Delta_X|}_{\text{確率変数}} \geq \underbrace{|\Delta_x|}_{\text{観測値}} \mid \underbrace{\{\mathcal{O}_X, \mathcal{M}_X\} = \{\mathcal{O}_x, \mathcal{B}_x\}}_{\text{注目領域と背景領域が観測値のものと同じ}} \right) \\ &= \Pr \left(\underbrace{|\Delta_X|}_{\text{確率変数}} \geq \underbrace{|\Delta_x|}_{\text{観測値}} \mid \underbrace{X_i \geq \theta, i \in \mathcal{O}_x, X_i < \theta, i \in \mathcal{B}_x}_{\text{注目領域と背景領域が観測値のものと同じ}} \right) \end{aligned}$$

特徴選択における条件付推論

- ▶ (通常の) 統計的有意性指標

$$p = \Pr \left(\underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{Y}},j}|}_{\text{確率変数}} \geq \underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{y}},j}|}_{\text{観測値}} \right)$$

- ▶ 条件付推論の統計的有意性指標

$$p = \Pr \left(\underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{Y}},j}|}_{\text{確率変数}} \geq \underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{y}},j}|}_{\text{観測値}} \mid \underbrace{\mathcal{M}_{\mathbf{Y}} = \mathcal{M}_{\mathbf{y}}}_{\text{選択された特徴集合が観測値のものと同じ}} \right)$$

特徴選択における条件付推論の例

- ▶ マージナルスクリーニング: $\mathbf{Y} \in \mathbb{R}^n$ と $\mathbf{x}_j \in \mathbb{R}^n$ の相関の大きな特徴を k 個選択

$$\underbrace{\mathbf{x}_{(1)}^\top \mathbf{Y} \geq \mathbf{x}_{(2)}^\top \mathbf{Y} \geq \mathbf{x}_{(3)}^\top \mathbf{Y}}_{\text{selected (when } k=3\text{)}} \geq \underbrace{\mathbf{x}_{(4)}^\top \mathbf{Y} \geq \mathbf{x}_{(5)}^\top \mathbf{Y} \geq \dots \geq \mathbf{x}_{(D)}^\top \mathbf{Y}}_{\text{not selected (when } k=3\text{)}},$$

(注) 各変数が正規化されているとすると相関は内積

- ▶ 仮説選択イベント

$$\begin{array}{ccc} \mathbf{x}_{(1)}^\top \mathbf{Y} \geq \mathbf{x}_{(4)}^\top \mathbf{Y} & \mathbf{x}_{(2)}^\top \mathbf{Y} \geq \mathbf{x}_{(4)}^\top \mathbf{Y} & \mathbf{x}_{(3)}^\top \mathbf{Y} \geq \mathbf{x}_{(4)}^\top \mathbf{Y} \\ \vdots & \vdots & \vdots \\ \mathbf{x}_{(1)}^\top \mathbf{Y} \geq \mathbf{x}_{(D)}^\top \mathbf{Y} & \mathbf{x}_{(2)}^\top \mathbf{Y} \geq \mathbf{x}_{(D)}^\top \mathbf{Y} & \mathbf{x}_{(3)}^\top \mathbf{Y} \geq \mathbf{x}_{(D)}^\top \mathbf{Y}, \end{array}$$

- ▶ 条件付推論の統計的有意性指標

$$\begin{aligned} p &= \Pr \left(\underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{Y}},j}|}_{\text{確率変数}} \geq \underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{y}},j}|}_{\text{観測値}} \mid \underbrace{\mathcal{M}_{\mathbf{Y}} = \mathcal{M}_{\mathbf{y}}}_{\text{選択された特徴集合が観測値のものと同じ}} \right) \\ &= \Pr \left(\underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{Y}},j}|}_{\text{確率変数}} \geq \underbrace{|\hat{\beta}_{\mathcal{M}_{\mathbf{y}},j}|}_{\text{観測値}} \mid \underbrace{\left\{ \mathbf{x}_{(\ell)}^\top \mathbf{Y} \geq \mathbf{x}_{(m)}^\top \mathbf{Y} \right\}_{(\ell,m) \in \{1,\dots,k\} \times \{k+1,\dots,d\}}}_{\text{選択された特徴集合が観測値のものと同じ}} \right) \end{aligned}$$

仮説選択イベント

セグメンテーションの場合

仮説選択イベント

$$X_i \geq \theta \text{ if } x_i \in \mathcal{O}_x, X_i < \theta \text{ if } x_i \in \mathcal{B}_x$$

特徴選択の場合

仮説選択イベント

$$\{\mathbf{x}_{(\ell)}^\top \mathbf{Y} \geq \mathbf{x}_{(m)}^\top \mathbf{Y}\}_{(\ell, m) \in \{1, \dots, k\} \times \{k+1, \dots, d\}}$$

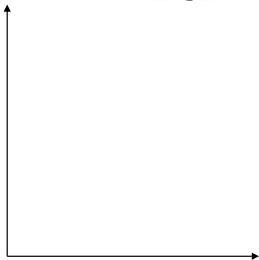
仮説選択イベント

セグメンテーションの場合

仮説選択イベント

$$X_i \geq \theta \text{ if } x_i \in \mathcal{O}_x, X_i < \theta \text{ if } x_i \in \mathcal{B}_x$$

データ空間 $X \in \mathbb{R}^n$

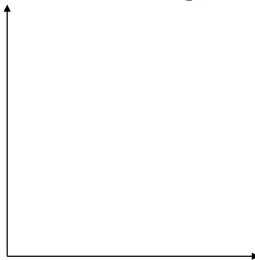


特徴選択の場合

仮説選択イベント

$$\{\mathbf{x}_{(\ell)}^\top \mathbf{Y} \geq \mathbf{x}_{(m)}^\top \mathbf{Y}\}_{(\ell, m) \in \{1, \dots, k\} \times \{k+1, \dots, d\}}$$

データ空間 $Y \in \mathbb{R}^n$



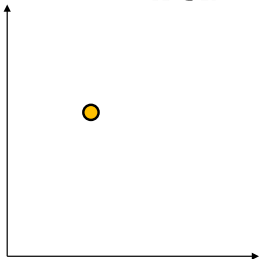
仮説選択イベント

セグメンテーションの場合

仮説選択イベント

$$X_i \geq \theta \text{ if } x_i \in \mathcal{O}_x, X_i < \theta \text{ if } x_i \in \mathcal{B}_x$$

データ空間 $X \in \mathbb{R}^n$

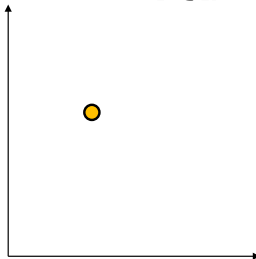


特徴選択の場合

仮説選択イベント

$$\{\mathbf{x}_{(\ell)}^\top \mathbf{Y} \geq \mathbf{x}_{(m)}^\top \mathbf{Y}\}_{(\ell, m) \in \{1, \dots, k\} \times \{k+1, \dots, d\}}$$

データ空間 $Y \in \mathbb{R}^n$

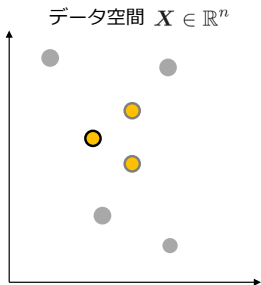


仮説選択イベント

セグメンテーションの場合

仮説選択イベント

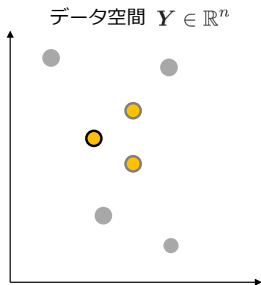
$$X_i \geq \theta \text{ if } x_i \in \mathcal{O}_x, X_i < \theta \text{ if } x_i \in \mathcal{B}_x$$



特徴選択の場合

仮説選択イベント

$$\{\mathbf{x}_{(\ell)}^\top \mathbf{Y} \geq \mathbf{x}_{(m)}^\top \mathbf{Y}\}_{(\ell, m) \in \{1, \dots, k\} \times \{k+1, \dots, d\}}$$

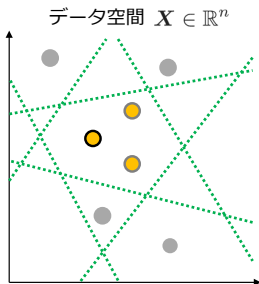


仮説選択イベント

セグメンテーションの場合

仮説選択イベント

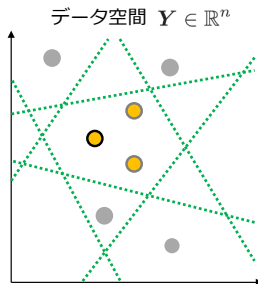
$$X_i \geq \theta \text{ if } x_i \in \mathcal{O}_x, X_i < \theta \text{ if } x_i \in \mathcal{B}_x$$



特徴選択の場合

仮説選択イベント

$$\{\mathbf{x}_{(\ell)}^\top \mathbf{Y} \geq \mathbf{x}_{(m)}^\top \mathbf{Y}\}_{(\ell, m) \in \{1, \dots, k\} \times \{k+1, \dots, d\}}$$

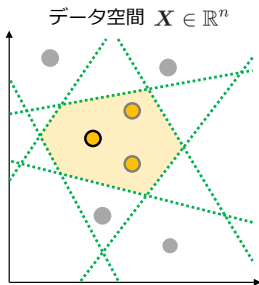


仮説選択イベント

セグメンテーションの場合

仮説選択イベント

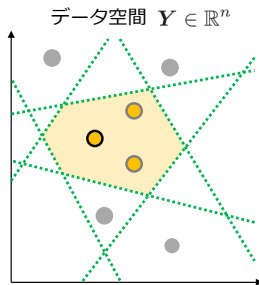
$$X_i \geq \theta \text{ if } x_i \in \mathcal{O}_x, X_i < \theta \text{ if } x_i \in \mathcal{B}_x$$



特徴選択の場合

仮説選択イベント

$$\{\mathbf{x}_{(\ell)}^\top \mathbf{Y} \geq \mathbf{x}_{(m)}^\top \mathbf{Y}\}_{(\ell,m) \in \{1,\dots,k\} \times \{k+1,\dots,d\}}$$



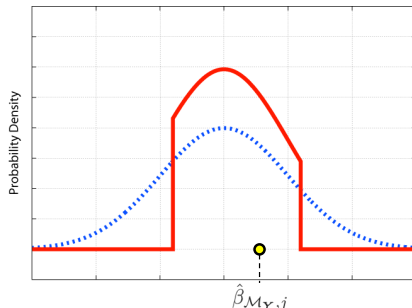
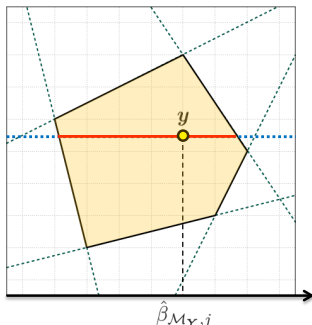
多面体の補題 (Lee+16)

- ▶ (Lee+2016) の研究が、ポスト選択推論の研究が進化するきっかけとなった
- ▶ (Lee+2016) は、Lasso により選択された特徴の条件付選択的推論を提案
- ▶ 多面体の補題 (Polyhedral Lemma) の骨子

条件付選択的推論において、2つの条件：

1. 仮説選択イベントがデータ空間の多面体（線形不等式の集合）
2. 検定統計量がデータの線形関数

を満たせば、**切断正規分布に基づいて正確で適切な p 値 (exact valid p -value)** を計算できる



Lasso により選択された特徴の条件付選択的推論 (Lee+16)

- ▶ Lasso を、データから特徴と符号を選択するアルゴリズムとみなす：

$$\mathcal{A}^{\text{Lasso}} : \mathbf{Y} \mapsto \{\mathcal{M}, \mathbf{s}\}$$

ただし、 $\mathcal{M}$ は選択された特徴集合、 $\mathbf{s}$ はそれらの符号のベクトル

- ▶ 選択された特徴の検定統計量 (再掲)

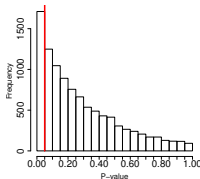
$$\hat{\beta}_{\mathcal{M}_{\mathbf{Y}},j} = (X_{\mathcal{M}_{\mathbf{Y}}}^{\top} X_{\mathcal{M}_{\mathbf{Y}}})^{-1} X_{\mathcal{M}_{\mathbf{Y}}}^{\top} \mathbf{Y} = \boldsymbol{\eta}^{\top} \mathbf{Y}$$

- ▶ 選択的 p 値 (selective p -value)

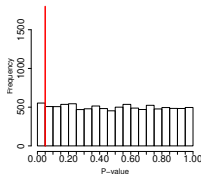
$$p = \Pr \left(|\hat{\beta}_{\mathcal{M}_{\mathbf{Y}},j}| \geq |\hat{\beta}_{\mathcal{M}_{\mathbf{y}},j}| \mid \underbrace{\mathcal{M}_{\mathbf{Y}} = \mathcal{M}_{\mathbf{y}}}_{\text{特徴の集合}}, \underbrace{\mathbf{s}_{\mathbf{Y}} = \mathbf{s}_{\mathbf{y}}}_{\text{特徴の符号}}, \underbrace{P_{\boldsymbol{\eta}}^{\perp} \mathbf{Y} = P_{\boldsymbol{\eta}}^{\perp} \mathbf{y}}_{\text{独立な方向}} \right)$$

- ▶ 選択的 p 値は exact で valid

$$\Pr_{H_0} (p \leq \alpha \mid \mathcal{M}_{\mathbf{Y}} = \mathcal{M}_{\mathbf{y}}, \mathbf{s}_{\mathbf{Y}} = \mathbf{s}_{\mathbf{y}}) = \alpha \quad \forall \alpha \in (0, 1)$$



H_0 における通常の p 値の分布



H_0 における選択的 p 値の分布

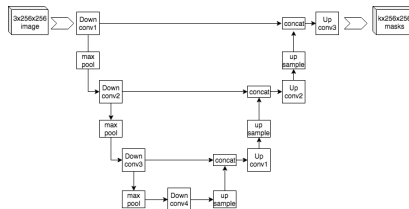
Part 2 のまとめ

- ▶ 条件付推論は選択的推論の有望なアプローチの一つとして注目上昇中
- ▶ 仮説選択イベントが多面体で表されれば正確で適切な p 値の計算が可能
- ▶ 仮説選択イベントはアルゴリズム依存 — 複雑なアルゴリズムでの適用可能性は未知

Part 3: 深層学習モデルのポスト選択推論

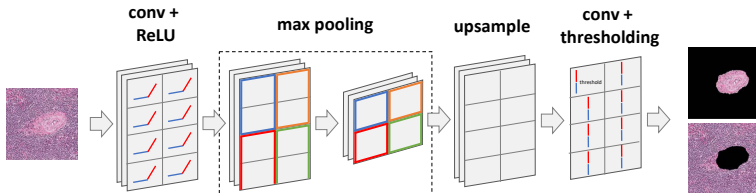
深層学習モデルによる画像セグメンテーション

► U-net (CNN)



U-Netの基本構造 (wikipediaページより)

► CNN の基本要素



深層学習モデルの仮説選択イベント

- ▶ CNN に基づくセグメンテーションアルゴリズム

$$\mathcal{A}^{\text{CNN}} : \mathbf{X} \mapsto \{\mathcal{O}_{\mathbf{X}}, \mathcal{B}_{\mathbf{X}}\}$$

- ▶ 条件付選択的推論

$$p = \Pr \left(|\Delta_{\mathbf{X}}| \geq |\Delta_{\mathbf{x}}| \mid \underbrace{\mathcal{A}^{\text{CNN}}(\mathbf{X}) = \mathcal{A}^{\text{CNN}}(\mathbf{x})}_{\text{CNN の出力が観測値のものと同じ}}, \underbrace{P_{\eta}^{\perp} \mathbf{X} = P_{\eta}^{\perp} \mathbf{x}}_{\text{独立な方向}} \right)$$

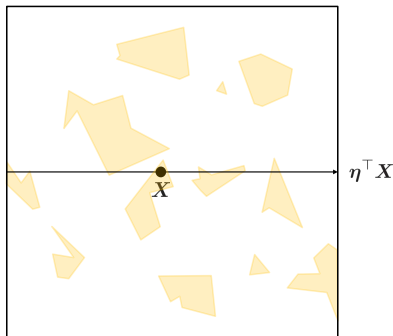
- ▶ Q. CNN の仮説選択イベント

$$\{\mathcal{O}_{\mathbf{X}}, \mathcal{B}_{\mathbf{X}}\} = \{\mathcal{O}_{\mathbf{x}}, \mathcal{B}_{\mathbf{x}}\} \Leftrightarrow \mathcal{A}^{\text{CNN}}(\mathbf{X}) = \mathcal{A}^{\text{CNN}}(\mathbf{x})$$

を定式化して計算することができるか？

分割統治アプローチ

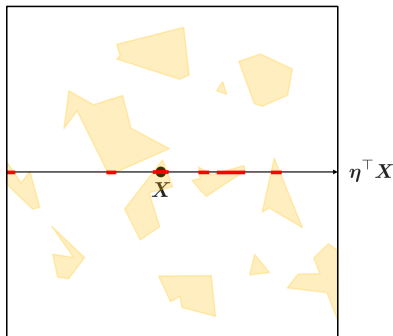
- ▶ step 1: 問題を過剰条件イベントに分割し、それぞれの切断正規分布を求める
- ▶ step 2: 複数の切断正規分布を統合して統計的推測を行う



過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチ

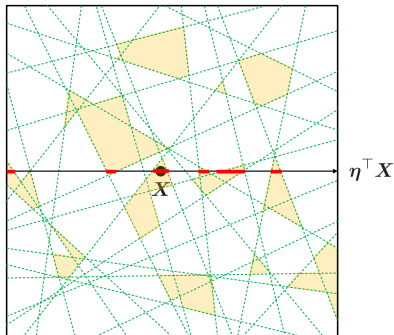
- ▶ step 1: 問題を過剰条件イベントに分割し、それぞれの切断正規分布を求める
- ▶ step 2: 複数の切断正規分布を統合して統計的推測を行う



過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチ

- ▶ step 1: 問題を過剰条件イベントに分割し、それぞれの切断正規分布を求める
- ▶ step 2: 複数の切断正規分布を統合して統計的推測を行う

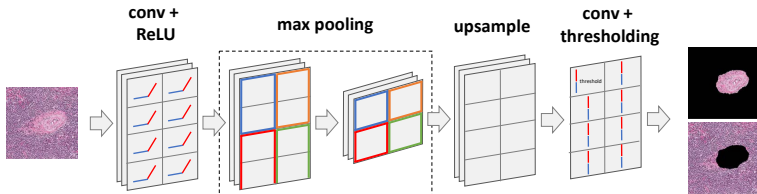


過剰条件イベントの統合によるポスト選択推論の概念図

CNN の基本要素

▶ CNN の基本要素の演算

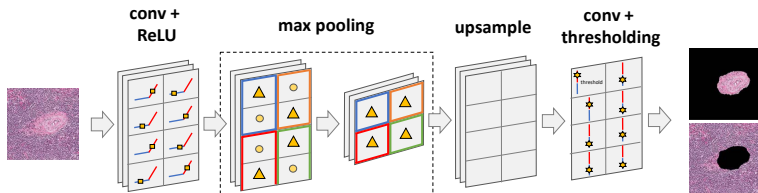
要素	演算
Convolution	線形
ReLU transfer function	区分線形
Max-pooling	大小比較
Upsampling	線形
Thresholding	閾値比較



CNN の基本要素

▶ CNN の基本要素の演算

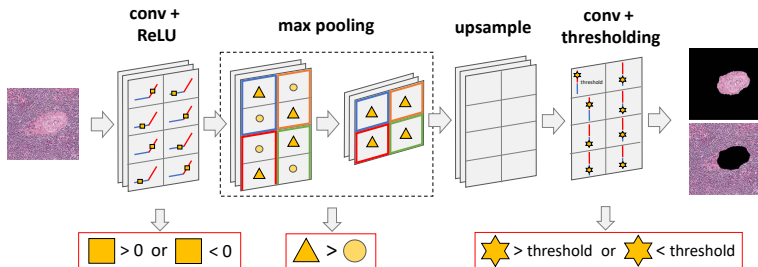
要素	演算
Convolution	線形
ReLU transfer function	区分線形
Max-pooling	大小比較
Upsampling	線形
Thresholding	閾値比較



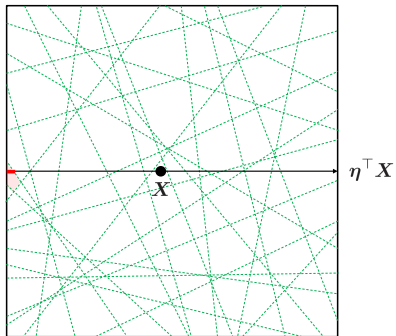
CNN の基本要素

▶ CNN の基本要素の演算

要素	演算
Convolution	線形
ReLU transfer function	区分線形
Max-pooling	大小比較
Upsampling	線形
Thresholding	閾値比較

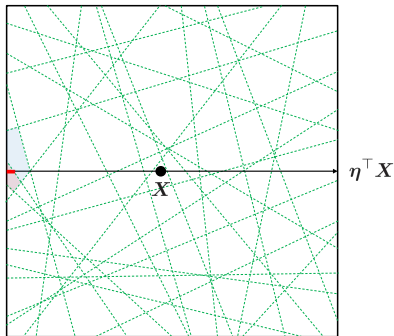


分割統治アプローチの概要



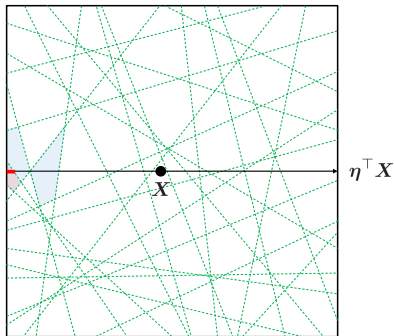
過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチの概要



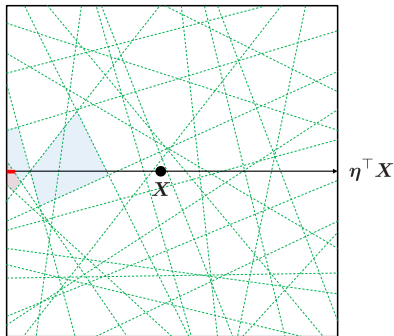
過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチの概要



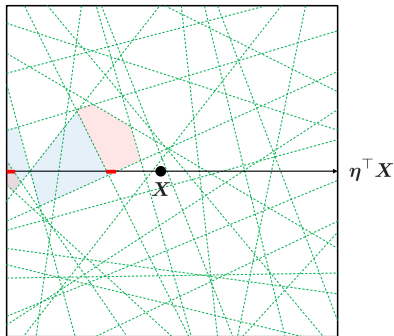
過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチの概要



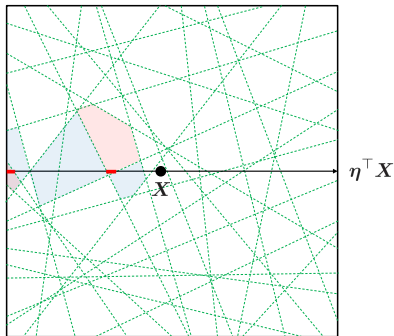
過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチの概要



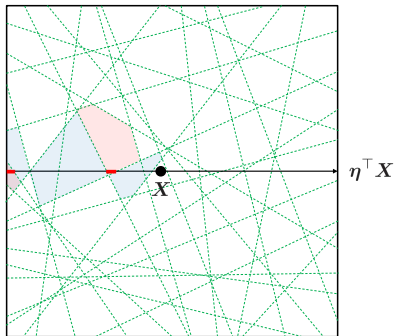
過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチの概要



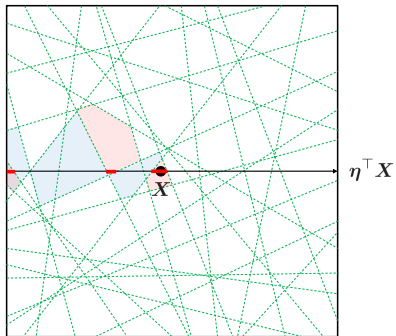
過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチの概要



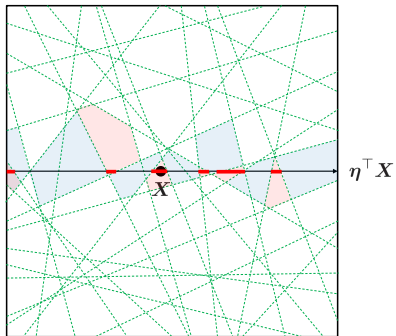
過剰条件イベントの統合によるポスト選択推論の概念図

分割統治アプローチの概要



過剰条件イベントの統合によるポスト選択推論の概念図

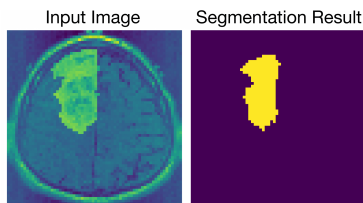
分割統治アプローチの概要



過剰条件イベントの統合によるポスト選択推論の概念図

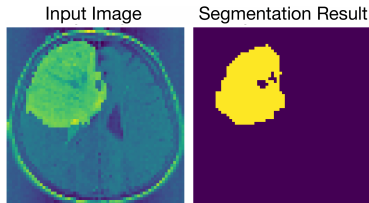
脳画像分析例

▶ 真に異常領域が存在する場合



通常の p 値： $\underbrace{0.000}_{\text{T.P.}}$

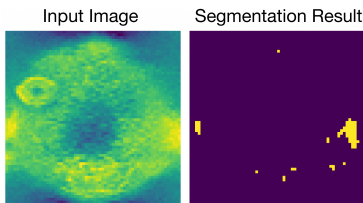
選択的 p 値： $\underbrace{0.000}_{\text{T.P.}}$



通常の p 値： $\underbrace{0.000}_{\text{T.P.}}$

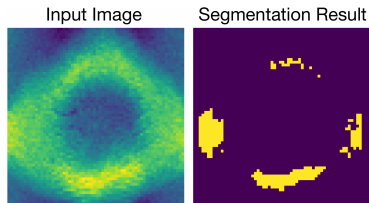
選択的 p 値： $\underbrace{0.000}_{\text{T.P.}}$

▶ 真に異常領域が存在しない場合



通常の p 値： $\underbrace{0.000}_{\text{F.P.}}$

選択的 p 値： $\underbrace{0.670}_{\text{T.N.}}$



選択的 p 値： $\underbrace{0.000}_{\text{F.P.}}$

新たな p 値： $\underbrace{0.451}_{\text{T.N.}}$

まとめ

- ▶ AI により発見される仮説はデータ駆動仮説であり、従来の統計解析では信頼性を正しく評価できない
- ▶ AI が産出する膨大な候補から最適な仮説が選ばれていると解釈すると、超多重比較問題とみなせる
- ▶ 多重比較では選択バイアスが生じるため補正が必要であるが、超多重比較問題における補正は難しい
- ▶ 既存の多重比較補正法に加え、条件付推論に基づく選択バイアス補正法が新たに注目を集めつつある
- ▶ 条件付推論はある仮説を出力するデータの集合を求める問題（アルゴリズムの逆問題）と解釈できる
- ▶ 条件付推論はアルゴリズム依存であり、アルゴリズムの柔軟性に応じて適切な補正を行うことが可能

- ▶ J. Lee et al. Exact post-selection inference, with application to the lasso. *Annals of Statistics* 2016.
- ▶ R. Tibshirani et al. Exact post-selection inference for sequential regression procedures. *Journal of American Statistical Association* 2016.
- ▶ W. Fithian et al. Optimal inference after model selection. *arXiv* 2014.
- ▶ J. Taylor et al. Statistical learning and selective inference. *PNAS* 2015.
- ▶ X. Tian, J. Taylor. Asymptotics of selective inference. *arXiv* 2015.
- ▶ S. Suzumura et al. Selective inference for sparse higher-order interaction models. *ICML2017*.
- ▶ K. Tanizaki et al. Computing valid p-values for image segmentation by selective inference. *CVPR2020*.
- ▶ VNL. Duy et al. Computing valid p-value for optimal changepoint by selective inference using dynamic programming. *NeurIPS2020*.
- ▶ VNL. Duy et al. Parametric programming approach for more powerful and general Lasso selective inference. *AISTATS2021*.
- ▶ VNL. Duy et al. Quantifying Statistical Significance of Neural Network-based Image Segmentation by Selective Inference. *arXiv* 2020.
- ▶ K. Sugiyam et al, More powerful and general selective inference for stepwise feature selection using homotopy method. *ICML2021*.
- ▶ D. Das et al. Fast and more powerful selective inference for sparse high-order interaction model. *AAAI2022*.