

# 神経画像解析における統計学的課題と アプローチ：確率場理論

内田 航<sup>1,2</sup>・青木 茂樹<sup>1,2</sup>・岩崎 学<sup>1,2</sup>

(受付 2025 年 7 月 1 日；改訂 8 月 23 日；採択 8 月 26 日)

## 要 旨

磁気共鳴画像法(MRI)に代表される神経画像技術の発展は、脳構造・機能の非侵襲的評価を可能にし、神経科学および臨床医学に飛躍的な進歩をもたらしてきた。特に、全脳探索的に脳の構造・機能的変化を統計学的に評価する手法は解剖学的・病理学的仮説に基づかない疾患評価を可能にし、疾患病態評価法として医学的にインパクトの高い知見が蓄積しつつある。しかし、神経画像データは本質的に高次元であり、全脳探索的解析における統計的推論では深刻な多重比較問題が生じる。また、隣接するボクセルで観測される MRI 特徴量間には複雑な空間的相関構造が存在し、古典的多重比較補正とは異なるアプローチの展開が求められてきた。本総説は、神経画像解析分野に特徴的な統計学的課題と、神経画像分野において長年応用されてきた確率場理論、現在の標準となっている統計学的手法の発展を概観する。

キーワード：神経画像, 多重比較, 確率場理論, permutation test, threshold-free cluster enhancement.

## 1. 神経画像解析における統計学的課題

磁気共鳴画像法(Magnetic Resonance Imaging; MRI)技術の目覚ましい発展は、ヒトの脳構造および機能を非侵襲的に三次元画像として捉えることを可能にし、神経科学研究や精神・神経疾患の病態解明に革新的な進歩をもたらした。特に、構造的 MRI (sMRI) は脳の形態学的特徴(例: 灰白質容積, 皮質厚などの巨視的な形態変化)の定量化を、拡散 MRI はその信号値の生物物理学的モデルに基づく神経微細構造の定量化, 機能的 MRI (fMRI) は特定の精神活動や刺激呈示に伴う脳血流動態変化の時空間的マッピングを可能にする。これらの情報は、アルツハイマー病, パーキンソン病などの神経変性疾患 (Uchida et al., 2023; Andica et al., 2023), 統合失調症, うつ病, 自閉症などの精神疾患といった (Nagai et al., 2024), 多様な疾患の早期診断マーカー開発や病態基盤の解明に加え、健常脳機能の理解深化に不可欠なものとなっている。例えば、本稿の後半(6.2 節)で紹介するタウオパチー研究のように、拡散 MRI を用いて神経線維の密度や太さといった微細構造を定量化し、臨床症状が類似する疾患群間で脳内のどこに違いが生じるのかを全脳にわたり探索的に比較することが可能となった (Uchida et al., 2023)。こうした探索的アプローチは、疾患の病態解明や客観的な診断マーカー開発への貢献が期待さ

<sup>1</sup> 順天堂大学 健康データサイエンス学部：〒279-0013 千葉県浦安市日の出 6 丁目 8-1; w-uchida@juntendo.ac.jp, saoki@juntendo.ac.jp, m.iwasaki.vs@juntendo.ac.jp

<sup>2</sup> 順天堂大学 大学院医学研究科：〒113-8421 東京都文京区本郷 2 丁目 1-1

れる。

一方で、これらの神経画像データから科学的に妥当な知見を抽出するためには、統計学的に特有かつ深刻な課題を克服する必要がある。まず、MRI データは本質的に極めて高次元である。例えば、Human connectome project で公開されている拡散 MRI では、1.25mm 立方のボクセル (voxel; 3次元画像の最小単位) 解像度で脳全体を撮像し、 $145 \times 174 \times 145 = 3,658,350$  個もの観測点が得られている (Glasser et al., 2013)。fMRI では、これに加えて時間軸方向の観測が数百点加わり、データはさらに高次元化する。神経・精神疾患を対象とした神経画像研究においては、一般に全脳のどこに脳の機能的・構造的変化が存在するのか自明でない場合が多く、探索的な病態評価の側面が強い。この場合、全脳ボクセルに渡る網羅的な統計学的評価が重要になるが、先の特徴から多重比較の問題が常態化していると言えよう。古典的には、関心のある疾患において仮説に基づいて脳の生物学的に意味のある局所的な構造を Region of interest (ROI) として定義し、局所領域における平均的な特徴量を抽出することで、この問題に対処してきた。しかし、このアプローチでは事前に定義した ROI 内の特徴量の平均(または中央値など)を算出するため、探索的な観点からは不十分であり、ボクセルレベルの探索的アプローチが求められる。以上のように、MRI のような大規模かつ複雑なデータ構造に対して探索的に統計的推論を行う際、各ボクセルで仮説検定を同時に実行することに伴う多重比較問題が深刻な課題となるのである。

MRI 画像に特徴的なもう一つの性質に、各ボクセルにおける信号値間には複雑な空間的相関構造が存在する点が挙げられる。これは、脳組織の解剖学的連続性により隣接する細胞が類似した構造・機能特性を持つこと、撮像時の点拡がり関数による空間的平滑化効果などに起因する。この空間的相関の存在下では、従来の多重比較補正法は適切に機能せず、統計的推論の妥当性が著しく損なわれる可能性がある。また、解剖学的にも、その変化は単一のボクセルではなくある程度の空間的広がりを持つクラスターとして現れることが期待されるため、クラスターレベルでの推論が有効となる。

本稿では、神経画像データにおいて、統計学的に考慮すべき特徴と派生する課題、その課題に対処すべく発展してきた統計学的手法の応用について概説する。

## 2. 神経画像分野における統計モデリングの基本と前処理

### 2.1 一般線形モデル (General Linear Model)

神経画像解析における全脳探索的解析においては、各ボクセル  $k$  において一般線形モデル (General Linear Model; GLM) が広く用いられる。 $n$  人の被験者からなるデータセットにおいて、ボクセル  $k$  での観測値を要素とするベクトルを  $\mathbf{Y}_k$  とすると、GLM は以下のように定式化される：

$$(2.1) \quad \mathbf{Y}_k = \mathbf{X}\boldsymbol{\beta}_k + \boldsymbol{\epsilon}_k, \quad k = 1, 2, \dots, V$$

ここで、

- $\mathbf{Y}_k$ : ボクセル  $k$  における  $n \times 1$  の観測データベクトル (MRI 特徴量)。
- $\mathbf{X}$ :  $n \times q$  の計画行列 (design matrix)。各列が説明変数 (例: 群のダミー変数, 年齢, 性別など) を表し、全てのボクセルで共通に用いられることが多い。群のダミー変数は、例えば、ある疾患群と健常対照群の間で各ボクセルの特徴量に差があるかを検証する場合に設定される。
- $\boldsymbol{\beta}_k$ : ボクセル  $k$  における  $q \times 1$  の未知の回帰係数ベクトル。これが推定対象となる。
- $\boldsymbol{\epsilon}_k$ :  $n \times 1$  の誤差ベクトル。

誤差ベクトル  $\epsilon_k$  の各要素は、互いに独立に平均 0、分散  $\sigma^2$  の正規分布に従うと仮定される。パラメータ  $\beta_k$  は通常、最小二乗法(または重み付き最小二乗法)によって推定される： $\hat{\beta}_k = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}_k$ 。特定の仮説を検証するためには、コントラストベクトル  $\mathbf{c}$ (例えば、2 群間比較の場合、 $[1, -1, 0, \dots, 0]^T$ )を用いて、 $\mathbf{c}^T \beta_k$  の値を評価する。帰無仮説  $H_0: \mathbf{c}^T \beta_k = 0$  の下で、検定統計量(例：t 統計量や F 統計量)が各ボクセルで計算される。例えば、t 統計量は、

$$(2.2) \quad T_k = \frac{\mathbf{c}^T \hat{\beta}_k}{\sqrt{\widehat{\text{Var}}(\mathbf{c}^T \hat{\beta}_k)}} = \frac{\mathbf{c}^T \hat{\beta}_k}{\sqrt{\hat{\sigma}_k^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}}$$

として得られ、帰無仮説の下で自由度  $n - q$  の t 分布に近似的に従う。添字  $k$  はボクセルのインデックスを表し、 $T_k$  の値は脳内の全ボクセルについて個別に算出される。各ボクセルで算出された  $T_k$  の値を空間座標場に配置した統計マップが、後続の多重比較補正の対象となる。なお、第 4 章で述べる確率場理論では、この統計マップを確率場  $Z(\mathbf{x})$  の実現値として扱う。

## 2.2 前処理の統計学的意義と影響

生の MRI データは、ノイズやアーチファクトを多く含み、また被験者間で脳の形状やサイズ、位置が異なるため、統計解析を行う前に一連の前処理が必要となる。これらの前処理ステップは、後続の統計解析の妥当性や感度に大きな影響を与える。本稿では、今日まで神経画像解析において広く利用されている Statistical Parametric Mapping (SPM) (Friston, 2003) における一般的な前処理についてまとめる。SPM は神経画像解析に必要な標準的な前処理に加え、統計解析までの一連の解析を実行可能なソフトウェアであり、今日まで広く利用されている。MRI 画像でしばしば見られるアーチファクトは、装置特性や撮像原理に依存して生じる。例えば、磁場不均一に起因する感度ムラや、高周波成分の打ち切りに起因する Gibbs アーチファクトなどがあり、種々補正アプローチが提案されているが (Tustison et al., 2010; Kellner et al., 2016)、SPM には実装されていない。実際に解析する MRI 画像に見られるアーチファクトに応じて個別に対処する必要があるが、本稿では詳細な方法論についての説明は割愛する。

### 2.2.1 空間標準化(Spatial Normalization)

異なる被験者の脳画像を共通の座標系(標準脳テンプレート、例えば MNI 座標系)に位置合わせする処理である。これにより、特定の脳部位に対応するボクセル座標が被験者間で概ね一致し、群レベルでの全脳探索的な統計解析が可能になる。標準化は、通常、線形変換(平行移動、回転、スケーリング、せん断)と非線形変換(局所的な変形)を組み合わせで行われる。標準化の精度は、特に小領域の解析において結果の信頼性に直結し、ボクセル毎の検定を行う場合は 1 ボクセルを構成する数 mm 以内の誤差を担保する必要があるが、技術的に難しく、SPM では次項の空間平滑化で対処する。

### 2.2.2 空間平滑化(Spatial Smoothing)

画像に対してガウシアンカーネルなどの空間フィルタを適用し、意図的に画像をぼかす処理である。平滑化の主な目的は以下の通りである：

- (1) 信号対雑音比の向上：ランダムノイズを平均化することで、信号の検出感度を上げる。
- (2) データ分布の正規性改善：中心極限定理に似た効果により、データの分布を正規分布に近づける。これは確率場理論(Random Field Theory, 以下 RFT) のガウス性仮定を満たす助けとなる。
- (3) 空間標準化の残存誤差の緩和：標準化後も残る微小な解剖学的位置ずれの影響を軽減し、群内での対応付けを改善する。

(4) **RFT** の仮定充足：RFT が適用可能となるために必要なデータの平滑性を確保する。

しかし、過度な平滑化は空間解像度を低下させ、小さな活性化領域を見逃す原因となる。平滑化カーネルのサイズ選択は、期待される効果の空間的広がりとのトレードオフの関係にあり、結果に大きな影響を与える重要なパラメータである。そのため、目的となる脳構造により最適化が必要であるが、一般に全脳探索的な統計解析を実施する場合には、慣習的に半値全幅(full width at half maximum, 以下 FWHM) = 8mm 程度に設定される場合が多い。

### 3. 神経画像解析における多重比較問題

神経画像解析における全脳探索的研究では、脳内の各ボクセルにおいて、特定の効果の有無に関する仮説検定を同時に行う。例えば、ある疾患群と健常対照群の間で脳の局所的な灰白質容積に差があるかを検証する場合、数万オーダーのボクセルそれぞれに対して帰無仮説の検定を実施することになる。仮に、各ボクセルでの検定を有意水準  $\alpha = 0.05$ で行うと仮定する。もし全ての帰無仮説が真であるとしても、1回の検定で偽陽性を犯す確率は  $\alpha$  であるため、全ボクセル  $V$  個での偽陽性の期待数は、 $E[\text{偽陽性数}] = V \times \alpha$  となる。前述の  $V = 145 \times 174 \times 145$  の場合、 $3,679,375 \times 0.05 \approx 183,969$  個となり、到底容認できない数の偽陽性が期待される。これは、神経画像研究の信頼性を著しく損なう。この問題に対処するため、多重比較補正が実施され、第 1 種の過誤確率(Family-Wise Error Rate; FWER)の制御が頻用される。

#### 3.1 Bonferroni 補正とその保守性

最も単純な FWER 制御法は Bonferroni 補正である。これは、個々の検定の有意水準を  $\alpha_{\text{voxel}} = \alpha_{\text{FWER}}/V$  と設定するものである。これらの手法は、検定統計量がどのような依存構造を持っていても(あるいは独立であっても)FWER を  $\alpha_{\text{FWER}}$  以下に制御するが、検定統計量間に正の相関が存在する場合には極めて保守的となる。脳画像データでは、隣接ボクセル間の統計量は生物学的・物理学的要因から強く正に相関している点を考慮すると、実効的に独立な検定の数は実際のボクセル数よりもはるかに少ないはずである。そのため、Bonferroni 補正を適用すると、個々のボクセルに課される有意水準が過度に厳しくなり、真の効果を見逃す第二種の過誤確率が増大し、検出力が著しく低下する。空間相関を考慮する試みの一つとして、実効的な独立検定数を推定し、Bonferroni 補正を行う方法 (Nyholt, 2004)があるが、安定した推定が困難であり、神経画像領域では広く利用されるに至っていない。

#### 3.2 ファミリーワイズエラー率 (Family-Wise Error Rate; FWER)

FWER は、検定のファミリー(ここでは全  $V$  個の検定)において、少なくとも 1つの真の帰無仮説を誤って棄却する確率を指す。ここで、前述の一般線形モデルに基づく検定統計量を  $T_k$  とし、ある閾値  $u$  を超える場合は帰無仮説  $\mathcal{H}_0$  を棄却するとする。この時、FWER は下記で表される。

$$(3.1) \quad \text{FWER} = Pr \left( \bigcup_{k \in \mathcal{H}_0} \{|T_k| \geq u\} \right) = Pr \left( \max_{k \in \mathcal{H}_0} |T_k| \geq u \right)$$

ここで  $\mathcal{H}_0$  は真の帰無仮説の集合(効果が実際に存在しないボクセルのインデックス集合)、 $u$  は棄却のための閾値である。すなわち、FWER は帰無仮説  $\mathcal{H}_0$  の下で、検定統計量  $T_k$  の最大値分布として表される。これは、閾値  $u$  を超えるクラスターが一つ以上存在することと同値である。次章に示す RFT では、この最大値分布を求める。

#### 4. 確率場理論(Random Field Theory, RFT)

神経画像データの持つ空間的相関構造を考慮し、Bonferroni 補正の保守性を克服する試みとして、1990 年代に RFT が導入された (Worsley et al., 1992; Friston et al., 1994a, 1994b). RFT は、ボクセルごとの統計値 (t 値, F 値など) からなるマップを連続的な確率場 (Random Field) の離散的な実現値とみなし、その確率場のトポロジカルな特性 (特にオイラー標数) を利用して、閾値を超える領域の最大値分布やクラスターサイズの分布を導出することで、空間的相関を考慮した FWER 制御を行う。この手法は SPM などの主要な解析ソフトウェアパッケージに実装され、長らく神経画像統計解析の標準的手法として広く用いられてきた。

##### 4.1 ガウス確率場とオイラー標数

神経画像解析領域においては、ボクセルごとの統計量がガウス確率場 (Gaussian Random Field; GRF) に従うという仮定に依存している。GRF  $Z(\mathbf{x}), \mathbf{x} \in S \subset \mathbb{R}^D$  (ここで  $D$  は空間次元、神経画像では通常 3 次元) は、探索空間  $S$  (例: 全脳領域) 内の任意の位置  $\mathbf{x}$  (個々のボクセル位置に対応) における統計値  $Z(\mathbf{x})$  が、どの有限個の点の組  $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_m))$  をとっても、それらが多変量正規分布に従うような確率場である。

RFT では、この確率場がある高い閾値  $u$  を超える領域、すなわち excursion 集合  $A_u = \{\mathbf{x} \in S : Z(\mathbf{x}) \geq u\}$  の幾何学的・トポロジカルな特性を利用する。特に重要なのがオイラー標数 (Euler Characteristic; EC)  $\chi(A_u)$  である。3 次元の場合、オイラー標数は、

$$(4.1) \quad \chi(A_u) = (\text{連結成分の数}) - (\text{穴の数}) + (\text{空洞の数})$$

と定義される。神経画像解析の文脈では、非常に高い閾値  $u$  に対しては、「穴」や「空洞」は稀であり、 $\chi(A_u)$  は閾値  $u$  を超えるボクセル群が形成する連結した塊、すなわちクラスターの数に近似できる。図 1 では、例としてある 2 次元確率場と 3 つの異なる閾値を適応した場合のクラスターとオイラー標数の変化を示している。図 1(a) において、 $x$  軸と  $y$  軸に空間的座標が対応し、 $z$  軸方向に検定統計量が示されている。図 1(b) のように、ある閾値  $u$  におけるオイラー標数  $\chi_u$  は閾値  $u$  に依存して変化している様子が見て取れる。2 次元の場合、閾値を超える白い連結成分の数から、その中に存在する黒い領域 (穴) の数を引くことでオイラー標数が計算される。例えば、図 1(b) の  $Z > 0.18$  においては、連結した白いクラスタの数は 6 であり、穴の数は 3 であるため、オイラー標数は  $6 - 3 = 3$  となる。3 次元の場合も同様に、連結した塊の数から、ドーナツのようなトンネル状の構造 (穴) の数と、球殻のような内部に閉じられた空間 (空洞) の数を加減することで計算される。

Worsley ら (Worsley et al., 1992, 1996) により、GRF のオイラー標数の期待値  $E[\chi(A_u)]$  は、閾値  $u$ 、探索空間  $S$  の体積と形状、そして確率場の平滑性 (smoothness) の関数として解析的に与えられることが示された。実用上、この期待値は、探索空間の体積  $|S|$  と、画像の平滑性を示す RESEL (Resolution Elements) カウント  $R$  を用いて表現されることが多い。3 次元の神経画像の場合：

$$(4.2) \quad E\chi(A_u) \approx \frac{R(4 \ln 2)^{3/2}}{(2\pi)^2} (u^2 - 1) \exp\left(-\frac{u^2}{2}\right)$$

RESEL カウントは、 $R = \frac{|S|}{(\text{FWHM}_x \text{FWHM}_y \text{FWHM}_z)}$  で定義され、画像の各方向の半値全幅 (FWHM) を考慮した実効的な空間解像度の単位である。データの空間的平滑性の推定には、一般線形モデル (GLM) の残差の空間勾配から推定されるのが一般的である (Kiebel et al., 1999)。この推定の精度は RFT の妥当性に直接影響するため、適切な平滑化処理と正確な平滑性推定が重要となる。

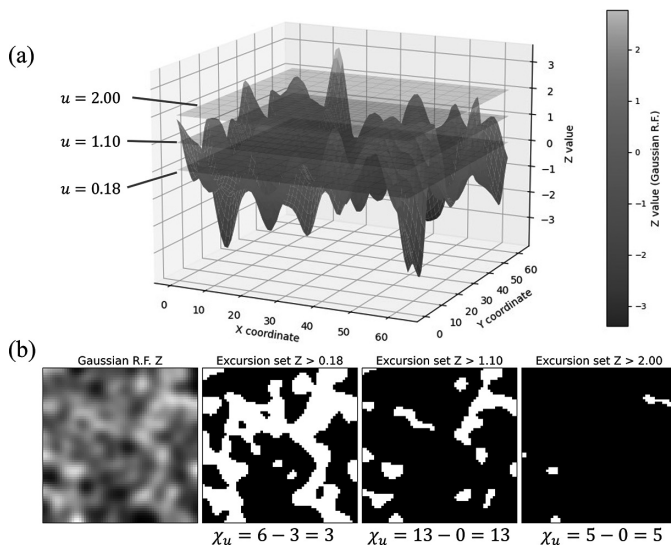


図 1. (a)生成された 2 次元ガウス確率場の 3 次元サーフェスプロット．図中には、3 つの異なる閾値 ( $u = 0.18, 1.10, 2.00$ ) が半透明な水平平面として重ねて表示されている．カラーバーは確率場の Z 値のスケールを示す．(b)左端の図は、(a)で示された元の 2 次元ガウス確率場 (Gaussian R.F. Z) をグレースケールで表示したヒートマップ．続く 3 つの図は、それぞれ閾値  $u = 0.18, 1.10, 2.00$  における excursion 集合 (確率場の値 Z が閾値  $u$  を超える領域、すなわち  $Z > u$ ) を二値画像で示したものである．これらの二値画像では、閾値を超える領域が白で、それ以外の領域が黒で表示されている．各 excursion 集合の下部には、その領域の連結成分数から穴の数を引いて計算されたオイラー標数 ( $\chi_u$ ) が示されている．

#### 4.2 ボクセルレベル FWER 制御 (Peak Height Inference)

RFT によれば、帰無仮説 (すなわち、コントラストベクトル  $c$  で定義される効果  $\theta$  が、全てのボクセルで一様にゼロであるという仮説) の下で、確率場の最大値  $Z_{\max} = \max_{\mathbf{x} \in S} Z(\mathbf{x})$  がある閾値  $u$  を超える確率  $Pr(Z_{\max} \geq u)$  は、高い  $u$  に対して  $E[\chi(A_u)]$  で近似できる：

$$(4.3) \quad \text{FWER}_{\text{voxel}} = Pr(Z_{\max} \geq u) \approx E[\chi(A_u)]$$

神経画像解析では、 $Z_{\max}$  は全脳で観測された統計値マップ中の最大値 (ピークの高さ) に対応する．この近似を用いて、指定した FWER 水準  $\alpha$  (例：0.05) を与える閾値  $u_\alpha$  を逆算することができる．個々のボクセルの統計値がこの  $u_\alpha$  を超えた場合に、そのボクセルは統計的に有意であると判定する．

#### 4.3 クラスターレベル FWER 制御 (Cluster Extent Inference)

ボクセルレベルの推論は、個々のボクセルでの信号強度が非常に高い場合に有効だが、空間的に広がった比較的弱い信号の検出には不向きである．そこで、RFT はクラスターレベルの推論も提供する．これは以下の手順で行われる：

- (1) まず、個々のボクセルで検定統計量  $T_k$  を計算する．
- (2) 次に、任意のクラスター定義閾値  $u_{\text{CDT}}$  (CDT, Cluster-Defining Threshold) (例：対応する片側 p 値が 0.001 や 0.01 となるような統計値) を設定する．

- (3)  $T_k \geq u_{\text{CDT}}$  を満たすボクセルを特定し、隣接する(例：SPM では 18 連結で定義)有意ボクセルを連結してクラスターを形成する。
- (4) 形成された各クラスターについて、そのサイズ(体積またはボクセル数)  $S$  を計算する。
- (5) 帰無仮説の下で、サイズ  $s$  以上のクラスターが偶然生じる確率  $Pr(S_{\text{max}} \geq s)$  を RFT に基づいて評価し、FWER 制御された  $p$  値をクラスターごとに算出する。ここで  $S_{\text{max}}$  は観測されうる最大のクラスターサイズである。

クラスターサイズ  $S$  の分布に関して、Friston et al. (1994a) は、帰無仮説のもとで観測される最大のクラスターサイズが  $s$  以上である確率  $Pr(S_{\text{max}} > s)$  は、

$$(4.4) \quad P(S_{\text{max}} \geq s) = 1 - \exp[-E[m] \cdot e^{-\beta s^{2/D}}]$$

と近似されることを示した。ここで  $E[m]$  は CDT の設定による閾値  $u_{\text{CDT}}$  を超えるクラスターの期待数(これも  $E[\chi(A_{u_{\text{CDT}}})]$  から導かれる)、 $e^{-\beta s^{2/D}}$  はある特定のクラスターのサイズが  $s$  を超える条件付き確率であり、近似的に  $\Pr(-\beta s^{2/D})$ 、 $\beta$  は  $\beta = [\Gamma(D/2 + 1) \cdot E[m]/E[N]]^{2/D}$  ( $D$  は空間次元、 $E[N]$  は閾値を超えるボクセルの期待数)の形で与えられる。この理論に基づき、観測されたクラスターサイズに対する補正済み  $p$  値が計算される。

## 5. RFT の限界とノンパラメトリック多重比較補正への展開

RFT は神経画像解析における多重比較補正の標準的手法として長年用いられてきたが、その理論的妥当性はいくつかの仮定に依存している。RFT の主な仮定には、(1)統計量が(少なくとも裾の部分で)ガウス確率場(または関連する確率場)として適切に近似できること、(2)確率場が連続かつ十分に滑らかであること(一般に FWHM がボクセルサイズの 2-3 倍以上)、(3)確率場の相関構造が脳全体で均一であること、などがある。これらの仮定、特にガウス性と平滑性の仮定は、実データでは必ずしも満たされない。また、クラスターを定義するための閾値(CDT)設定の恣意性も問題となる。このような背景から、データの分布に関する仮定をほとんど置かないノンパラメトリックなアプローチ、CDT 設定の恣意性を克服するアプローチを組み合わせた手法が注目されるようになった。代表的なものとして、Permutation Test と、それと組み合わせる用いられることが多い閾値フリークラスター強調(Threshold-Free Cluster Enhancement; TFCE)がある。ここでは概要に留めるが、これらの手法は、近年の神経画像コミュニティにおいて標準的な解析手法の一つとなっている。

### 5.1 Permutation Test

Permutation test は、R. A. Fisher によって提唱されたノンパラメトリックな統計的推論手法であり、検定統計量の帰無分布に関して特定の確率分布を仮定しない(Fisher, 1935; Nichols and Holmes, 2002; Winkler et al., 2014)。具体的には、群ラベルをランダムに並べ替える操作を多数回行い、各データセットに対して検定統計量を計算することで、帰無仮説の下での検定統計量の経験的な帰無分布を構築する。多重比較補正に適用する際は、各データセットにおいて「ファミリー全体で最も極端な統計量」(例えば、全ボクセル中の最大  $t$  値や、後述する最大 TFCE 値)を求め、その分布を構築する(Westfall and Young, 1993)。観測データから得られた統計値が、この最大統計量の帰無分布の上位  $\alpha \times 100$  パーセンタイル点を超えるかどうかで有意性を判断する。この手順により、データの分布形状や空間相関構造について明示的な仮定を置くことなく、FWER を制御できる。年齢や性別など複数の共変量を含む場合、単純な群ラベルの並べ替えでは説明変数の対応が崩れるため、異なる手法が用いられる(Winkler et al., 2014)。詳細は割愛するが、まず共変量のみで説明される効果を回帰により取り除き、その残

り(残差)を被験者間でランダムに並べ替える。このアプローチにより、共変量の影響を適切に統制しながら、興味のある特定の変数(例えば群のダミー変数)の効果を頑健に検定することが可能となる。

## 5.2 Threshold-free cluster enhancement (TFCE) の概念

従来法では、クラスターを定義するための閾値(CDT)を研究者が任意に設定する必要があるが、CDT 閾値の選択によって結果が大きく変わりうるという問題があった(Woo et al., 2014)。閾値フリークラスター強調(TFCE)は、Smithら(Smith and Nichols, 2009)によって提案された、CDT の恣意性問題を克服するための手法である。

TFCE の基本的な考え方は、各ボクセルの元の統計値を、そのボクセルが属する全ての可能なクラスターの「高さ(元の統計値)」と「広がり(クラスターサイズ)」の情報を統合的に反映するように変換することである。具体的には、あるボクセルに対して、考える様々な閾値レベルでそのボクセルが含まれるクラスターの大きさとその閾値自体の高さとを掛け合わせ、それらの値を低い閾値からそのボクセルの統計値まで積分する。数式での詳細な定義は原著(Smith and Nichols, 2009)に譲るが、この処理の結果、TFCE マップでは、元の統計マップにおいて局所的に高い信号強度を持つだけでなく、ある程度の空間的連続性・広がりを持って存在する信号領域が特に高い値を示すようになる。すなわち、単に「狭く鋭いピーク」を持つ領域や、「低強度だが広範囲」な領域よりも、信号強度(高さ)と空間的広がり(裾野)の双方をバランス良く兼ね備えた構造が強調される。これにより、特定の CDT を事前に主観的に選択する必要なく、統計的に意味のあるクラスター様の信号を強調し、その後の統計的推論における検出感度を高めることが期待される。

TFCE によって得られた強調統計マップと Permutation test を組み合わせることで、従来法の課題を克服し、近年の神経画像コミュニティにおいて標準的な解析手法の一つとして広く受け入れられつつあり、多くの主要な解析ソフトウェア(例: FMRIB Software Library [FSL] など)に実装されている。ただし、計算コストが高いという課題があり、効率的な実装や計算資源が求められる。

## 6. 臨床応用研究事例

神経画像統計解析手法は、基礎研究における脳機能・構造の理解深化に貢献するだけでなく、実際の臨床現場における疾患の診断支援や治療効果の客観的評価など、医学的応用も進んでいる。本節では、その代表的な事例を紹介する。

### 6.1 アルツハイマー病早期診断支援システム VSRAD

アルツハイマー病(AD)は、早期発見と早期介入が進行抑制に重要とされる神経変性疾患である。VSRAD(Voxel-based Specific Regional analysis system for Atrophy Detection)は、確率場理論に基づく多重比較補正は適応されていないが、前述の SPM の技術的基盤に基づいて脳の萎縮の程度を客観的かつ定量的に評価し、AD の早期診断を支援するために日本で開発されたシステムとして臨床的に広く利用されている(Matsuda et al., 2012)。VSRAD は、その客観的な評価指標と比較的簡便な操作性から、日本国内において医薬品医療機器等法(旧薬事法)に基づく医療機器としての承認を受け、2015 年時点で約 3000 の医療機関で AD の診断補助ツールとして臨床応用されている(松田, 2016)。VSRAD は、医師による視覚的読影に加えて定量的な萎縮情報を提供することで、より客観性の高い AD 診断の一助となっている。その有用性は複数の研究で示されており、例えば、Waragai et al. (2014)は健常高齢者と比較した AD の

診断能において感度 84.1%，特異度 84.6% を報告している。

## 6.2 タウオパチー研究

臨床応用研究の一例として、Uchida et al. (2023) らによる、進行性核上性麻痺 (Progressive Supranuclear Palsy; PSP) および大脳皮質基底核症候群 (Corticobasal Syndrome; CBS) 患者における白質の微細構造変化に関する研究を挙げる。PSP, CBS はいずれもタウオパチー (脳内に異常なタウタンパク質が蓄積し、その神経毒性に関連する神経変性を病態に持つ神経変性疾患群の総称) に分類される神経変性疾患であり、臨床的には病態が重複するため鑑別が困難な場合があるため、両疾患の鑑別に有用な客観的マーカーの開発が試みられている (Ling and Macerollo, 2018)。この研究では、PSP 患者群、CBS 患者群、および健常対照群を対象に、拡散 MRI を用いて計算された神経細胞微細構造を反映する特徴量の全脳探索的解析 (Fixel-based analysis, FBA) を行った。FBA では神経微細構造の特徴量として線維密度 (FD), 線維径 (log-FC), 両者の積 (FDC) が抽出可能であり、従来法との比較として萎縮などの巨視的な変化を捉える Voxel-based morphometry (VBM) と古典的な微細構造指標である Fractional Anisotropy (FA) の探索的な比較も行った。VBM, FA の検定は前述の RFT に基づく。FBA の統計解析においては、FBA で得られた統計値マップに対し、TFCE を拡張した Connectivity-based Fixel Enhancement (CFE) (Raffelt et al., 2015) を適用することでクラスター様の信号を強調し、その後 Permutation Test を実行することで、ファミリーワイズエラー率 (FWER) を厳密に制御した全脳探索的な多重比較補正を行った。その結果、PSP 患者群では健常対照群と比較して、上小脳脚 (SCP) や皮質脊髄路 (CST) を中心に神経細胞構造特徴量の有意な低下が認められた。一方、CBS 患者群では運動野下の白質、CST, 上縦束 (SLF), 脳梁などでより広範な特徴量の低下が示された (図 2)。これらの変化は両疾患の病理学的な知見と一致した脳領域に対応しており、非侵襲的 MRI 検査を用いた探索的な統計画像解析により、両疾患に特徴的な神経微細構造変化を捉えられる可能性が示唆された。さらに、SCP における細胞損傷は病理研究で PSP により特徴的であるこ

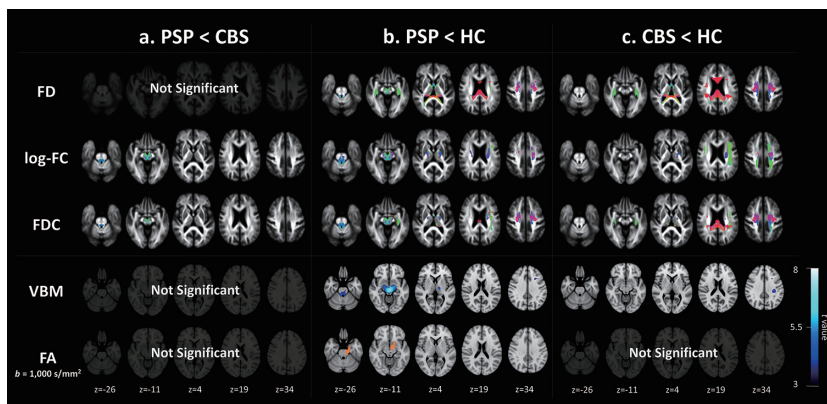


図 2. 進行性核上性麻痺 (PSP) と大脳皮質基底核症候群 (CBS) における白質微細構造変化の比較。(a) CBS と比較して PSP で有意に低値を示した領域, (b) 健常者 (HC) と比較して PSP で有意に低値を示した領域, (c) HC と比較して CBS で有意に低値を示した領域を示している (Uchida et al., 2023 より引用 (CC-BY))。図中上 3 つのパネルの有意領域の色は、神経線維が走行する方向に基づき RGB で示されている (赤: 左右方向, 青: 上下方向, 緑: 前後方向)。また、図中下 2 つのパネルの有意領域の色は、検定統計量  $t$  に基づいてカラースケール (最大値 8, 最小値 3) で示されている。

とが示されており, SCPにおけるMRI指標は臨床的に鑑別が難しいPSPとCBSの鑑別に寄与した。また, その変化は運動機能障害や認知機能障害といった臨床症状の重症度とも関連することが示された。この研究は, FBAとノンパラメトリックな統計手法を組み合わせることで, タウオパチーにおける疾患特異的かつ進行性の白質微細構造変化を感度良く捉え, 疾患の鑑別・予後予測マーカーとしてのその臨床的意義を示した例といえる。

## 7. まとめと展望

本総説では, 神経画像分野における統計解析手法, 特に多重比較問題への対応に焦点を当て, 古典的な確率場理論(RFT)から, 現代の標準となりつつあるTFCE-Permutation Testに至るまでの発展を概観した。神経画像統計解析の分野における統計学的アプローチは, 今後もさらなる発展が期待される。神経画像統計解析は, データ科学の一分野として, また現代神経科学を支える基盤技術として, 今後ますますその重要性を増していくであろう。本総説が, 当該分野への理解を深め, 統計学と神経科学のさらなる発展を促進するための一助となれば幸いである。

## 謝 辞

本研究は, 統計エキスパート人材育成プロジェクトの助成を受けた報告である。

## 参 考 文 献

- Andica, C., Kamagata, K., Uchida, W., Saito, Y., Takabayashi, K., Hagiwara, A., Takeshige-Amano, H., Hatano, T., Hattori, N. and Aoki, S. (2023). Fiber-specific white matter alterations in Parkinson's disease patients with GBA gene mutations, *Movement Disorders*, **38**(11), 2019–2030, <https://doi.org/10.1002/mds.29578>.
- Fisher, R. A. (1935). *The Design of Experiments*, Oliver & Boyd, Edinburgh.
- Friston, K. J. (2003). *Statistical Parametric Mapping*, 237–250, Springer US, Boston, Massachusetts, [https://doi.org/10.1007/978-1-4615-1079-6\\_16](https://doi.org/10.1007/978-1-4615-1079-6_16).
- Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C. and Evans, A. C. (1994a). Assessing the significance of focal activations using their spatial extent, *Human Brain Mapping*, **1**(3), 210–220, <https://doi.org/10.1002/hbm.460010306>.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D. and Frackowiak, R. S. (1994b). Statistical parametric maps in functional imaging: A general linear approach, *Human Brain Mapping*, **2**(4), 189–210, <https://doi.org/10.1002/hbm.460020402>.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R. et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project, *Neuroimage*, **80**, 105–124, <https://doi.org/10.1016/j.neuroimage.2013.04.127>.
- Kellner, E., Dhital, B., Kiselev, V. G. and Reiser, M. (2016). Gibbs-ringing artifact removal based on local subvoxel-shifts, *Magnetic Resonance in Medicine*, **76**(5), 1574–1581, <https://doi.org/10.1002/mrm.26054>.
- Kiebel, S. J., Poline, J.-B., Friston, K. J., Holmes, A. P. and Worsley, K. J. (1999). Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model, *NeuroImage*, **10**(6), 756–766, <https://doi.org/10.1006/nimg.1999.0508>.
- Ling, H. and Macerollo, A. (2018). Is it useful to classify PSP and CBD as different disorders? Yes, *Movement Disorders Clinical Practice*, **5**(2), 145–148, <https://doi.org/10.1002/mdc3.12581>.
- 松田博史 (2016). VSRAD®: 精神科臨床へのニューロイメージングの最新応用, *総合病院精神医学*, **28**(3), 210–218, <https://doi.org/10.11258/jjghp.28.210>.

- Matsuda, H., Mizumura, S., Nemoto, K., Yamashita, F., Imabayashi, E., Sato, N. and Asada, T. (2012). Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of probable Alzheimer disease, *American Journal of Neuroradiology*, **33**(6), 1109–1114, <https://doi.org/10.3174/ajnr.A2935>.
- Nagai, Y., Kirino, E., Tanaka, S., Usui, C., Inami, R., Inoue, R., Hattori, A., Uchida, W., Kamagata, K. and Aoki, S. (2024). Functional connectivity in autism spectrum disorder evaluated using rs-fMRI and DKI, *Cerebral Cortex*, **34**(13), 129–145, <https://doi.org/10.1093/cercor/bhad451>.
- Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples, *Human Brain Mapping*, **15**(1), 1–25, <https://doi.org/10.1002/hbm.1058>.
- Nyholt, D. R. (2004). A simple correction for multiple testing for SNPs in linkage disequilibrium with each other, *The American Journal of Human Genetics*, **74**(4), 765–769, <https://doi.org/10.1086/383251>.
- Raffelt, D. A., Smith, R. E., Ridgway, G. R., Tournier, J.-D., Vaughan, D. N., Rose, S., Henderson, R. and Connelly, A. (2015). Connectivity-based fixel enhancement: Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres, *Neuroimage*, **117**, 40–55, <https://doi.org/10.1016/j.neuroimage.2015.05.039>.
- Smith, S. M. and Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence, and localisation in cluster inference, *NeuroImage*, **44**(1), 83–98, <https://doi.org/10.1016/j.neuroimage.2008.03.061>.
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A. and Gee, J. C. (2010). N4ITK: Improved N3 bias correction, *IEEE Transactions on Medical Imaging*, **29**(6), 1310–1320, <https://doi.org/10.1109/TMI.2010.2046908>.
- Uchida, W., Kamagata, K., Andica, C., Takabayashi, K., Saito, Y., Owaki, M., Fujita, S., Hagiwara, A., Wada, A., Akashi, T. et al. (2023). Fiber-specific micro-and macroscopic white matter alterations in progressive supranuclear palsy and corticobasal syndrome, *npj Parkinson's Disease*, **9**(1), <https://doi.org/10.1038/s41531-023-00565-2>.
- Waragai, M., Hata, S., Suzuki, T., Mizukami, K., Asada, T. and Matsuda, H. (2014). Utility of SPM8 plus DARTEL (VSRAD) combined with magnetic resonance spectroscopy as adjunct techniques for screening and predicting dementia due to Alzheimer's disease in clinical practice, *Journal of Alzheimer's Disease*, **41**(4), 1207–1222, <https://doi.org/10.3233/JAD-132786>.
- Westfall, P. H. and Young, S. S. (1993). *Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment*, John Wiley & Sons, Edinburgh.
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. and Nichols, T. E. (2014). Permutation inference for the general linear model, *NeuroImage*, **92**, 381–397, <https://doi.org/10.1016/j.neuroimage.2014.01.060>.
- Woo, C.-W., Krishnan, A. and Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations, *NeuroImage*, **91**, 412–419, <https://doi.org/10.1016/j.neuroimage.2013.12.058>.
- Worsley, K. J., Evans, A. C., Marrett, S. and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain, *Journal of Cerebral Blood Flow & Metabolism*, **12**(6), 900–918, <https://doi.org/10.1038/jcbfm.1992.127>.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J. and Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation, *Human Brain Mapping*, **4**(1), 58–73, [https://doi.org/10.1002/\(SICI\)1097-0193\(1996\)4:1<58::AID-HBM4>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O).

## Challenges and Advances in Neuroimaging Statistical Analysis: Random Field Theory

Wataru Uchida<sup>1,2</sup>, Shigeki Aoki<sup>1,2</sup> and Manabu Iwasaki<sup>1,2</sup>

<sup>1</sup>Faculty of Health Data Science, Juntendo University

<sup>2</sup>Graduate School of Medicine, Juntendo University

The advancement of neuroimaging techniques, notably Magnetic Resonance Imaging (MRI), has enabled non-invasive assessment of brain structure and function, bringing remarkable progress to neuroscience and clinical medicine. In particular, statistical methods for whole-brain exploratory analysis of structural and functional brain changes have enabled disease evaluation independent of anatomical or pathological hypotheses, leading to the accumulation of medically impactful findings for disease pathophysiology assessment. However, neuroimaging data are inherently high-dimensional, creating a severe multiple comparisons problem in whole-brain exploratory statistical inference. Furthermore, complex spatial dependencies exist among MRI signals observed at the voxel level, necessitating the development of approaches different from classical multiple comparison corrections. This review provides an overview of the characteristic statistical challenges in neuroimaging analysis, the Random Field Theory that has been applied in the neuroimaging field for many years, and the evolution of current standard statistical methodologies.