

主論文の要旨

**Effect-Size Estimation Using Semiparametric  
Hierarchical Mixture Models in Disease-Association  
Studies with Neuroimaging Data**

脳画像を用いた疾患関連解析におけるセミパラメトリック階層  
混合モデルによる効果サイズ推定

名古屋大学大学院医学系研究科 総合医学専攻  
臨床医薬学講座 生物統計学分野

(指導：松井 茂之 教授)

江本 遼

## 【緒言】

認知症などの精神疾患は日本でも五大疾患に指定されており、社会の急速な高齢化に伴い、精神疾患のメカニズム解明や有効な診断法の開発が喫緊の課題となっている。これらの課題の解決に資する研究の一つとして、患者と健常人の脳画像データを比較することで疾患に関連する部位のスクリーニングを行う疾患関連解析がある。この解析における統計学的な課題は、仮説検定による疾患に関連する部位の検出と、各部位の関連の大きさ(効果サイズ)の推定である。

従来の統計的手法の多くは疾患関連部位の検出に重点を置いており、近年では脳機能の局在性を考慮するために確率場理論の確率モデルを用いるなどの提案がみられる。一方、効果サイズ推定に関しては、検出された部位の単純な推定量が多くの研究で報告されているが、このような推定量には効果サイズを実際より大きく推定してしまうという推定バイアスの問題が指摘されている。しかし、この問題に対しては、検証用の集団を別途用意するなどの簡易的な対処法が提案されるに留まっている。そこで、本研究では、脳画像データの特徴である空間的な従属構造を考慮した、より推定バイアスの小さな効果サイズの推定法を提案する。

## 【方法】

脳画像の構成単位である各ボクセルと疾患との関連の大きさ、すなわち、効果サイズとして、ボクセル画素値に関する疾患群と健常群での平均値の差を群内の標準偏差で割った標準化平均差  $\delta$  を用いるのが自然である。なお、実際には、 $\delta$  の推定量として、データから求めた平均値と標準誤差を用いた推定量  $Y$  を用いる。

統計量  $Y$  の統計学的モデルを考える際には、隣接するボクセル間には強い相関が存在するという従属構造、及び、効果サイズ分布の形状に関する情報が欠如していることをモデルに反映させる必要がある。本研究では、ボクセル間の従属構造をモデルに組み込む隠れマルコフ確率場モデルと、効果サイズの分布の柔軟なモデリングを可能にするセミパラメトリック階層混合モデルを統合する。

具体的には、任意のボクセル  $s$  に対して、これが疾患に関連しているかを表す潜在変数  $\theta_s$  ( $\theta_s = 1$  ならばボクセル  $s$  は疾患と関連あり、 $\theta_s = 0$  ならば関連なし) を導入し、全てのボクセルに関しての  $\theta_s$  の同時分布がマフコフ確率場モデルの一つであるイジングモデルに従うと仮定する。サンプルサイズが十分大きい場合、漸近正規性により、疾患と関連しないボクセルの推定量  $Y_s$  は平均  $\delta_s = 0$ 、分散  $c_n^2$  の正規分布に従う。すなわち、 $Y_s \sim N(0, c_n)$  である。ここで  $c_n$  はサンプルサイズによって定まる定数である。一方、疾患と関連するボクセルでは、特定の分布形を指定しないノンパラメトリック分布  $g$  を用いて、 $Y_s$  の分布は以下のセミパラメトリック階層混合モデルで表されると仮定する。

$$Y_s | \delta_s, \theta_s = 1 \sim N(\delta_s, c_n^2), \\ \delta_s \sim g(\cdot).$$

なお、以上の分布の仮定には正規近似(漸近正規性)を用いているが、サンプルサイ

ズが小さい場合は、正規近似を用いない  $t$  分布を用いたモデルも考慮する。

以上のモデルのもとでの疾患との関連の有無の仮説検定は local index of significance (LIS) 統計量に基づく偽発見率 (false discovery rate) の調整により行う。また、ノンパラメトリック分布  $g$  を離散分布で近似する手法を用いて効果サイズの経験ベイズ推定を考えると、効果サイズの推定量は事後期待値  $d_s = E[\delta_s | y_s, \Theta_s = 1; \hat{\varphi}]$  と事後確率  $\ell_s = \Pr(\Theta_s = 1 | y; \hat{\varphi})$  の積の形で  $\hat{\delta}_s = d_s \ell_s$  として導かれる。ここで、 $\varphi$  はモデルに含まれるパラメータであり、推定量  $\hat{\varphi}$  は一般化 EM アルゴリズムを用いて求めることができる。

### 【結果】

提案する効果サイズ推定量  $\hat{\delta}_s$  の真の値  $\delta_s$  からの乖離を、様々な設定のもと数値実験により確認した。サンプルサイズ及びイジングモデルのパラメータの値によらず、提案する推定量の推定バイアスは、従来の推定法よりも小さかった (Figure 1)。また、特にサンプルサイズが小さい場合に、 $t$  分布を用いる提案法の推定量は、正規近似を用いる推定量よりバイアスが小さかった。

実データへの適用として、Open Access Series of Imaging Studies (OASIS) における OASIS-2 という公開データ (Marcus et al., 2010) を用いて、経度認知障害 (MCI) からアルツハイマー病 (AD) への進行と脳の部位の関連解析を提案手法により行った。画像データの前処理として OASIS-2 で公開されている T1 強調画像をソフトウェア SPM により標準化、正規化し、各部位の容積を表す画像へ変換した後、提案手法を適用した (Figure 2)。Figure 2 に示すように、従来の仮説検定に重点を置いた方法と異なり、関連の有無だけでなく効果サイズの大きさを得ることができる。また、推定された効果サイズの大きい部位の多くは、従来の研究で関連が指摘されている領域であり、従来の研究とも一貫した結果であった (Table 1)。

### 【考察】

提案法により、多重検定を通じた疾患関連ボクセルの検出だけでなく、個々のボクセルのより正確な効果サイズの推定が可能となる。Figure 2 (b) に示すように、個々のボクセルの効果サイズの推定により、検出された脳領域と疾患との関連に関する詳細で偏りのない情報が提供でき、検出された領域の生物学的または臨床的検討に役立つことが期待される。また、検定統計量ではなく効果サイズの統計モデルを用いたことにより、サンプルサイズに依存しない評価が可能となっており、サンプルサイズの異なる研究間での効果サイズの比較に役立つことも期待される。さらに、効果サイズは患者背景などの共変量が存在する場合にも拡張可能であり、提案する推定の枠組みは、線形モデルを用いる fMRI や DTI データなど多くの脳画像を用いた関連解析に適用可能である。

提案した推定法では、正規近似を用いたモデルと  $t$  分布を用いたモデルそれぞれに対して効果サイズの推定量を得ることができる。数値実験による両者の推定性能の比

較では、サンプルサイズが小さい場合、正規近似を用いた推定量は $t$ 分布を用いた推定量に劣るが、サンプルサイズが増えるにつれて同程度になる傾向が認められた。正規近似の利用は計算時間短縮の観点から有用であるため、サンプルサイズが十分に大きい場合(例えば 100 例以上)は正規近似を用いた方法を使用することを推奨する。

最後に、本研究の重要な側面として、脳画像データに含まれる多数のボクセル全体を通しての効果サイズの分布を柔軟にモデリングしていることがあげられる。これにより、ある脳領域もしくは隣接するボクセルといったごく小さな領域に対しても正確な推論を行うことが可能となる。また、提案したモデルの拡張により脳の空間的構造に加えて機能的構造を考慮するといった発展も可能である。

### **【結語】**

本研究は脳画像を用いた疾患関連解析において、効果サイズの過大推定バイアスに対する有効な補正法の開発に取り組んだ最初の試みである。提案手法は、脳画像の従属構造を組み込んだ柔軟な階層セミパラメトリックモデルに基づいて、各ボクセルの効果サイズの正確な推定、および疾患に関連するボクセルの検出を可能にする。