

## 脳画像解析における統計学的クラスター推測

川口 淳

佐賀大学医学部地域医療科学教育研究センター

## はじめに

脳画像解析において VBM (Voxel Based Morphometry) に代表される全脳を対象とし、探索的に疾患関連部位を特定するためには、複数の被験者に対するデータから健常者に比べて罹患者のボクセル (画像の最小単位) 値が異なるかどうかを判断する。この判断の際には統計学に基づいた方法である仮説検定が頻用されている。データから算出された量 (ボクセル値の平均値の群間差) が意味のある大きさであることを、仮説検定を用いて客観的に判断するためには、帰無仮説 (差なし、もしくは関連なし) を仮定した上で規定されるその群間差の基準化量 (検定統計量) に対する分布を用いて、検定統計量がどれほど帰無仮説から乖離しているかを評価する必要がある。その評価量として p 値があり有意水準と呼ばれる閾値より p 値が小さくなれば帰無仮説を棄却して対立仮説が成り立つ、すなわち、“有意に” 差あり (もしくは関連あり) という統計学的な結論を得る。有意水準は多くの場合 0.05 が採用される。このような統計学的推測の背後には、結論に対する二種類の誤りが存在する。これは主に被験者の多様性によるデータの不確実性に基づくものである。仮説検定はそのうちの第 1 種の過誤 (もしくは偽陽性) と呼ばれる、帰無仮説が正しいのに帰無仮説を棄却してしまうという誤りを小さく抑えた上で判断を行う方法である。その誤りが起こる確率である過誤確率は、実は有

意水準と等しい、すなわち有意水準を 0.05 と設定した仮説検定では、第 1 種の過誤確率が 0.05 になることが保証されているのである。もう一つの誤りであり第 2 種の過誤 (もしくは偽陰性) と呼ばれる、対立仮説が正しいのに帰無仮説を棄却しない誤りは検定を行う上では考慮されていない。

脳画像解析では様々なデザインの研究に対応できる一般線形モデル (GLM, General Linear Model) が主な解析方法となっており、やはり仮説検定を行った上で解析結果が得られる。脳画像解析における GLM についての詳細は、文献<sup>1),2)</sup>を参照にされたい。GLM は脳画像の最小単位であるボクセルごとに適用され、全脳であれば数百万ボクセル回検定を繰り返している (多重検定) ことになり、上述の過誤確率が検定実施ごとに積み重なり、結果的に 0.05 より大きい過誤確率になってしまうので、多重補正を行った上で帰無仮説の棄却判断が必要となる。多重補正については文献<sup>3)</sup>を参照されたい。

元来、仮説検定は臨床試験など検証的な解析においてよく用いられており、多くの場合は単一の評価項目に対して用いられてきた。しかし、多面的な評価を行うためには複数の評価項目が必要となったことに伴い多重検定の方法論が検証的な場面を想定して様々発展しており、近年では脳画像や遺伝子解析のような膨大なデータを解析するための方法としても用いられている。全脳を対象としたボクセル単位で行う

この総説は、第 44 回日本磁気共鳴医学会大会緊急シンポジウム「クラスターレベルでの統計的推論の注意点～fMRI データ統計解析にまつわる最近の論争をめぐって～」での講演を中心にまとめたものである。

キーワード brain image, statistical hypothesis test, multiple comparison, cluster size test, random field theory

脳画像解析において多重検定を適用する際に問題となるのはその検定すべきボクセル数の多さであり、伝統的な多重補正法では有意水準が高くなりすぎる、すなわち、厳しすぎるがゆえに帰無仮説がなかなか棄却されず結論を得にくいという保守的な方法となってしまう。これを解決するために、検定数を減らすためにボクセルではなくボクセルの集まり、すなわち、クラスターを形成した上で検定を行う方法（クラスター推測）が発展し、脳画像統計解析の標準的方法として多くのソフトウェアで実行できるようになった。

脳画像でクラスターを考える際には、脳の構造上、距離的に近い近隣のボクセルは似たような値をもつとするのが自然であり、隣接するボクセルの集まりをクラスターとしてボクセルごとに有するデータの情報を解析する。このためにクラスターの推測（検定）を行う際にボクセル間の相関を考慮する必要があり、高度な数学の知識を要する確率場理論を利用する。ユーザーの多くは数学を非専門としているために、ブラックボックス的にソフトウェアを使用することもあるかもしれない。このような状況を危惧したのか、Eklund ら<sup>4)</sup>がセンセーショナルな見出しとともに警告のような実験結果を公表した。その見出しと昨今の情勢が後押しして誤解を生むような情報も飛び交うため、本論ではVBMの解析例を基にクラスター推測の解説をした上で、Eklund ら<sup>4)</sup>とその後の議論の続きなどをまとめた上で見解を述べる。

### クラスターレベル推測

本節ではこのクラスターを対象にした推測であるクラスターレベル推測を解説する。はじめにクラスターの作り方を解説し、各クラスターのサイズ（クラスター内ボクセル数）を検定する方法を、比較的ユーザーの多い解析ソフトウェア SPM (Statistical Parametric Mapping) の方法に基づき解説する。

#### 1. SPM の結果

本論で使用する SPM バージョン 12

(SPM12) を用いた GLM の適用例を示す。画像データは、ワシントン大学アルツハイマー病研究グループによる Open Access Series of Imaging Studies (OASIS, <http://www.oasis-brains.org/>) が公開しているデータベースから国際的評価法「臨床的認知症尺度」(CDR; 正常 (=0), 疑い (=0.5), 軽度 (=1), 中等度 (=2), 高度 (=3)) が測定されている被験者を選択した。また CDR  $\geq 1$  をアルツハイマー病 (AD) 疑いとし、CDR=0 を健常 (NC) とした。灰白質 T<sub>1</sub> 強調画像の品質が良い AD 群 22 名に対して、年齢、性別、脳内容積、教育歴をマッチングさせた NC 群 22 名を選択した。AD 群と NC 群の群間比較を GLM により行う。前処理として解剖学的標準化と平滑化を、SPM の拡張ツールボックスである CAT12 (Computational Anatomy Toolbox) により行った画像を解析した。ボクセルサイズは 1.5 mm 立方としたため処理後のボクセル数は 121 × 145 × 121 = 2,122,945 個であった。平滑化におけるパラメータである FWHM (Full Width at Half Maximum) は 8 mm にした。GLM の結果を Fig. 1 に示す。Fig. 1(A) はボクセルごとに行われた t 検定の結果として、ボクセルごとに算出された検定統計量 (T 値) が脳画像テンプレート上に図示される T-map を示している (閾値は任意設置のものである)。これを基にいくつかの数値を設定すると Fig. 1(B) のような統計解析結果の表が得られる。

#### 2. クラスター

全脳解析の GLM において Fig. 1(A) のようにボクセルごとに検定統計量 T 値が算出され作成された T-map に対するクラスターは Table 1 の手順で形成される。

①における閾値は height threshold, cluster-

Table 1. Cluster generation procedure

- |   |
|---|
| ①検定統計量に対する任意の閾値を決める。                                  |
| ②その閾値以上の値をもつボクセルのうち隣接するボクセルの塊を一つのクラスターと考え、クラスターを形成する。 |

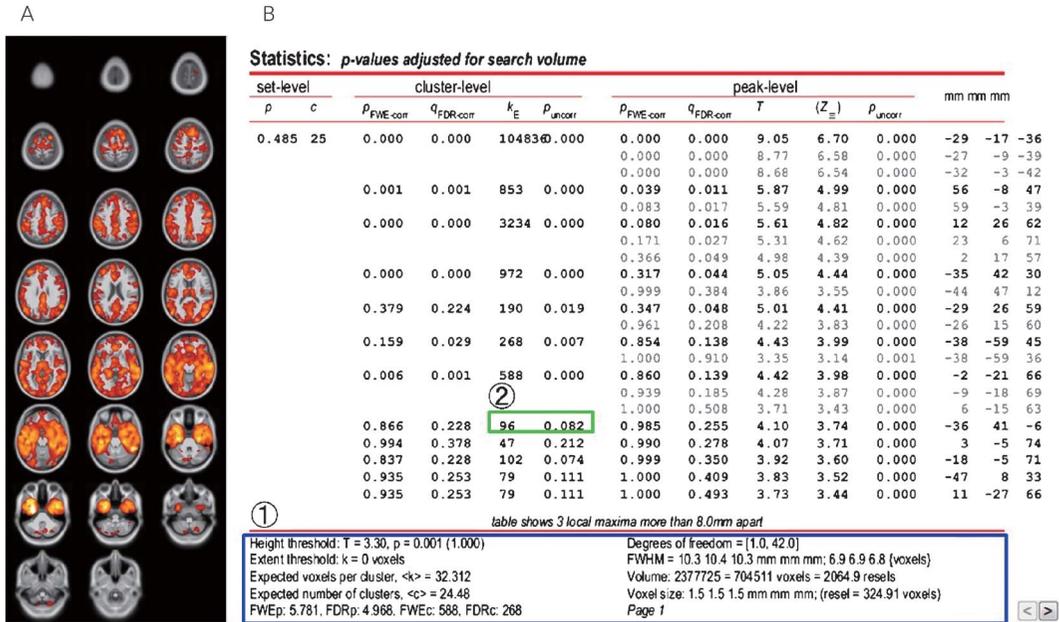


Fig. 1. Result for SPM (A) T-map (B) Statistics table

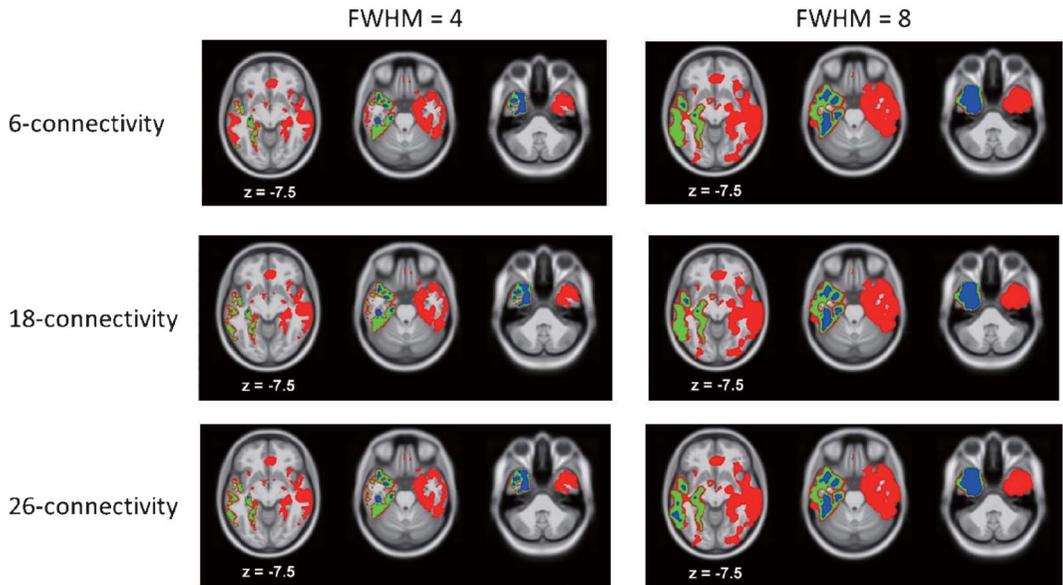


Fig. 2. Cluster generation result

defining threshold (CDT), cluster forming threshold などと呼ばれるが、本論では CDT と呼ぶことにする。これは検定統計量の大きさでもあるが、それに 1 対 1 で対応する p 値で与えられる場合もある。SPM のデフォルトは

uncorrected  $p = 0.001$  であり、これは 1 節の例では自由度 42 であり検定統計量で言えば  $T = 3.3$  に対応する。Fig. 1(B)の①部分の左列に height threshold として表示されている。matlab を使えば、Fig. 1(B)の①部分の右列にある

Degrees of freedom における 42 を自由度として,  $\text{tinv}(1-0.001, 42)$  と入力すると求めることができる。

②においてはどこまでのボクセルを“隣接”とするかという定義が必要になる。3次元に配置されたあるボクセルは同一の高さに左右前後斜めに8個のボクセルが隣接している。その真上に9個、真下に9個隣接している。これら全てを隣接と考えるのが 26-connectivity という方法である (8+9+9=26個のボクセルが隣接)。また、真上と真下に位置するボクセルのうち四隅のものを隣接しないと考える方法が 18-connectivity という方法である (26-4-4=18個のボクセルが隣接)。またそのあるボクセルに対して面の部分しか接していないボクセルを隣接とする方法が 6-connectivity である (立方体の面の数=6個のボクセルが隣接)。26, 18, 6-connectivity の順で大きいクラスターが形成されそうである。SPM では 18-connectivity を使用している。②においては検定を行う前の各被験者画像に対する前処理である平滑化が大きく影響する。平滑化によって隣り合うボクセル値がより近い値をもつようになる (画像が滑らかになる)。この際にどの程度を近い値にしていくかを FWHM というパラメータによって規定する。これは縦横高さ方向に3つの値が設定されるが、ほとんどが同じ値を設定するので、本論ではその共通する値一つで表す。1節では 8 mm とした。この値が高ければ原画像をより広い範囲のボクセルを近い値に変換していることになる。このように滑らかにされた画像を用いた検定から得られた T-map も滑らかになると考えられる。よって FWHM が高い場合、大きなクラスターが出来やすくなる。

Fig. 2 には Fig. 1(A) の T-map に対して、CDT としては 3.3, 4, 5 とし 6, 18, 26-connectivity におけるクラスター形成を示している。前処理における FWHM は 4 と 8 を比較した。

Fig. 2 には複数形成されたクラスターのうち最大のものの一つ表示している。青色が各 CDT すべてで共通するボクセルであり、緑色が CDT=3.3 と 4 に共通するボクセル、赤色

が CDT=3.3 だけのボクセルである。すなわち CDT が低いとクラスターが大きくなることがわかる。さらには同じ CDT で見比べると FMWH=8 の方が大きいクラスターが形成されていることがわかる。6, 18, 26-connectivity の大きな違いは見られなかった。

### 3. クラスターサイズの検定

脳画像全体で大小様々な複数のクラスターが得られることになるが、どのクラスター (ボクセル) が有意に大きいかを p 値に基づき判断する。クラスターサイズとは各クラスターを形成しているボクセルの数である。Fig. 1(B) の cluster level における  $k_E$  の値は CDT=3.3 で得られたいくつかのクラスターサイズを表している。その検定における帰無仮説は「クラスターサイズ=0」である。算出された p 値が有意水準よりも小さいことにより帰無仮説が棄却されれば、そのクラスターは有意に大きいと判断する。p 値算出式の導出には確率場理論が用いられる。その仮定として、検定統計量は脳画像上で滑らかである、がある。したがって前処理として平滑化を実施しておく必要がある。確率場に基づくクラスター推測については文献<sup>5)~7)</sup>を参照されたい。

クラスターサイズ  $s$  に対する p 値  $P_s$  は次のように近似することができる。

$$P_s = \exp\left(-b\left[\frac{s}{r}h(h^2-1)\right]^{2/3}\right) \dots\dots\dots (1)$$

ただし  $b$  は定数で  $b=0.5334942\dots$  である。 $h$  は CDT を表し、 $r$  は RESEL を表す。RESEL は 3 方向の FWHM の積として求められる。注意としてここでの FWHM は T-map に対するものであり、前処理で設定されたものとは異なり、GLM の残差によって求められる。Fig. 1(B) の①部分の右列にあるように、3 方向の FWHM は 6.9, 6.9, 6.8 と計算されてある。その 2 行下に、これの積として  $\text{resel} = 324.92$  voxels と表示されてあるこの値が  $r$  であり、以下では四捨五入して  $r=325$  として用いる。

(1) 式は比較的簡単な式であり単純な計算機によっても計算することが可能である。 $P_s < 0.05$  であれば有意水準 0.05 (5%) で、クラス

ターサイズ  $s$  をもつクラスターは有意に大きいと判断できる。

さらには式(1)を  $s$  について解くと次のような式を得る。

$$s = \frac{r [-\log P_s/b]^{-2/3}}{h(h^2-1)} \dots\dots\dots (2)$$

この式の  $P_s$  に有意水準 (例えば 0.05) を代入するとそれに対応するクラスターサイズが計算され、それは代入された有意水準に対するクラスターサイズの閾値 (サイズ閾値) であり、その閾値より大きいクラスターはその有意水準において有意と判断することができる。これらの式の導出に関しては数学的にはなるが興味のある読者は付録を参照されたい。

Fig. 1(A) の T-map に対するこの計算式の使用例を示す。  $b=0.53$ ,  $h=3.3$ ,  $r=325$  として、  $s=96$  の検定を行う。式(1)に値を代入すると、

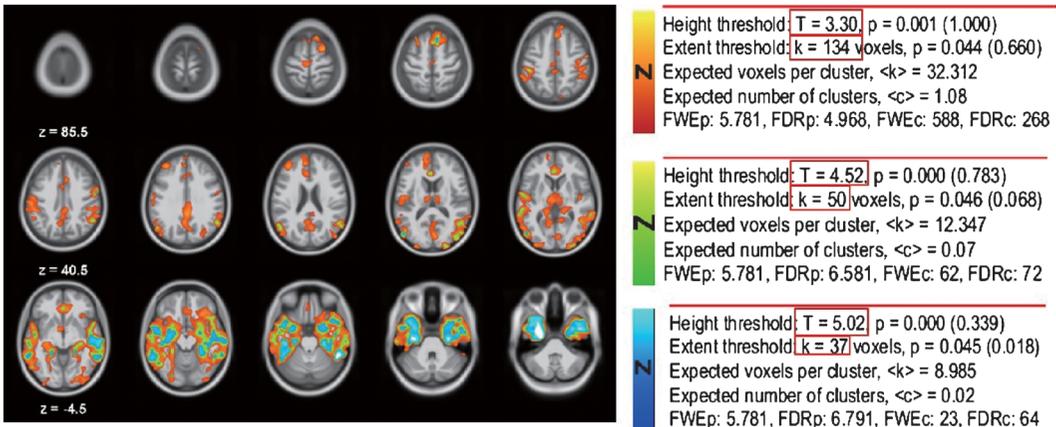
$$P_s = \exp \left( -0.53 \times \left[ \frac{96}{325} \times 3.3 \times (3.3^2 - 1) \right]^{2/3} \right) = 0.091$$

となり SPM 出力の 0.082 と近い値が算出されていることがわかる。再現性のために  $b$  の値を切り上げて計算しているが、 $b$  を正確な値で使用して計算すると  $P_s=0.089$  となりより SPM 出力に近づく。また、この状況下での有意水準 0.05 におけるサイズ閾値を求める。式(2)に値を代入すると、

$$s = \frac{325 \times [-\log 0.05/0.53]^{-2/3}}{3.3 \times (3.3^2 - 1)} = 133.8$$

これより 134 以上のクラスターサイズが有意水準 0.05 で有意ということになる。これは SPM の Extent threshold で 134 と入力することができる。Fig. 3 には Fig. 1 の T-map に対して、  $h=3.3$  に加え  $h=4.52$ ,  $5.02$  の場合 (それぞれ uncorrected  $p=0.000025$ ,  $0.000001$  に相当) に対しても式(2)から  $s$  を求め有意なクラスターを図示している。

Fig. 3 では Extent threshold 行の  $p$  の値が有意水準を示しており、ほとんど 0.05 と近い値が算出されており式(2)がよく近似されていることがわかる。閾値を設定するのは CAT12 の Transform SPM-maps という機能が有用である。



Degrees of freedom = [1.0, 42.0]  
 FWHM = 10.3 10.4 10.3 mm mm mm; 6.9 6.9 6.8 {voxels}  
 Volume: 2377725 = 704511 voxels = 2064.9 resels  
 Voxel size: 1.5 1.5 1.5 mm mm mm; (resel = 324.91 voxels)

Fig. 3. Significant cluster at 5% level (CDT=3.3, 4.5, 5)

4. クラスターサイズ検定の性質

3節において簡略された式(1)からクラスターサイズ検定のp値はCDTの $h$ とRESELの $r$ に依存することがわかる。この式の解釈において、これらの値は2節で見たようにクラスター形成にも影響するので、CDTまたはRESELが異なる値であれば検定されるクラスターサイズ(または形状)も異なっていることに注意しないといけない。

Fig. 4は同じクラスターサイズ $s=96$ に対して、 $h$ が変化するとp値およびサイズ閾値がどのように変化するかを図示したものである。また $r=125, 343, 729$ とした(それぞれFWHM=5, 7, 9に相当)。

$r$ が小さい方がFig. 4(A)からはp値は小さくなりそうであり、Fig. 4(B)からは閾値も低く設定できそうである。滑らかさが十分でなく作られたクラスターは、概して小さいものばかりであると考えられるので、例えば $r=125$ の場合の $s=96$ は比較的に大きいクラスターとみなされ、CDTに関係なく小さいp値を算出すると考えられる。このことを別の見方をする。式(1)において $s'=s/r$ として考えると $s'$ はRESELあたりのクラスターサイズと見ることができ。こうしてRESELに関して標準化されたクラスターサイズを評価するのが良いと考えられる。 $s'=s/r=96/343=0.28$ である。 $r=125$ の場合には、 $s'=0.28$ に相当するボクセル

数は $125 \times 0.28 = 35$ となる。こうして $s=96$ は $r=125$ では大きいクラスターと考えられる。こうしてみると滑らかでない方が検出力は高そうに見えるが、滑らかさが十分でなければ全体的にクラスターサイズが小さいクラスターしか得ることができず、結果の解釈に不具合が生じる可能性があるかもしれない。 $r$ は前処理の平滑化に大きく影響を受けるので、前処理FWHMの適切な値を設定する必要がある。これについては後に述べるが、SPMのデフォルトでもある8mmで良いと考える。

CDTが大きい方がFig. 4(A)からはp値は小さくなりそうであり、Fig. 4(B)からは閾値も低く設定できそうである。CDTを高くしてクラスターを作ったのであれば、サイズが小さくても有意になるとも解釈できる。CDTを低くするとクラスターはより大きなクラスターとして形成される可能性もあるが、その分p値が大きくなる傾向にある。閾値の方はCDT=3.0,  $r=343$ の時に200ほどである。CDTを小さくしすぎて大きなクラスターを形成してしまうと棄却されやすくなる。こうして、冒頭で述べたとおりに、この検定はクラスター形成の部分も考慮して解釈する必要がある。

5. 多重補正

これまでの節ではクラスター推論の基礎的な部分に着目し多重補正なしのp値を取り扱ったが、画像解析においては多重補正が必要とな

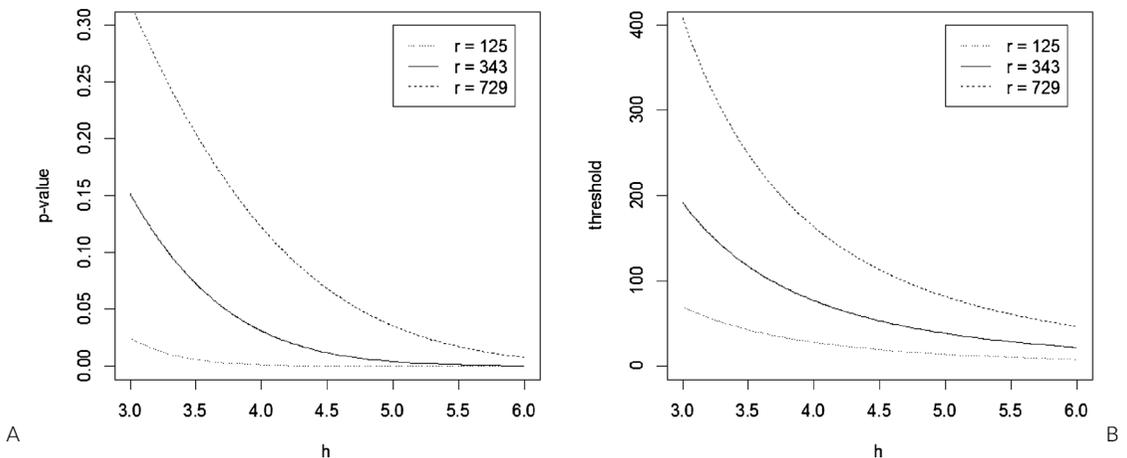


Fig. 4. p-value (A) and size threshold (B) for cluster size  $s=96$

る。SPM では FWE (Family Wise Error) と FDR (False Discovery Rate) の補正が可能である。他のソフトを用いると permutation test も可能である。多重補正の詳細については文献<sup>2),3)</sup>を参照されたい。

SPM の出力においては、Fig. 1(B)の①部分の左列最下行にある FWEc が FWE コントロールされたサイズ閾値であり、また FDRc は FDR コントロールされたものである。これらの値を、3 節で行ったように Extent threshold で入力することで多重補正を行ったことになる。Fig. 3 においてサイズ閾値を比較することができる。CDT=3.3 の場合には補正なしでは  $k=134$  であり、FWEc=588, FDRc=268 に対しては低いので偽陽性が高い可能性がある。実際に Extent threshold 行の ( ) 内には FWE の有意水準が示されており 0.660 と高い。これに対し CDT=4.5, 5 に対しては補正なしと FWEc と FDRc には大きな差がないと言える。CDT=5 となれば FWEc=23 となり、補正なしの  $k=37$  より保守的である。こうして FDRc による閾値がより良いように考えられる。一般的に多重補正は有意水準を低く設定することに対応するので、式(2)において有意水準を 0.01 にして計算し直すと、CDT=3.3, 4.5, 5 に対応するサイズ閾値は 253, 96, 69 と計算され、FDRc と近い値になることがわかる。さらなる詳細な検討が必要にはなるが、クラスター推測においては FDR 補正もしくは低い有意水準の uncorrected p が有効かもしれない。

### クラスター推測に対する論争

近年は Eklund ら<sup>4)</sup>の出版を皮切りに fMRI 解析におけるソフトウェア使用に対して議論が行われてきた。特に Eklund ら<sup>4)</sup>の偽陽性割合が 70% というセンセーショナルな数字によって騒然となった。ここで偽陽性割合 (false positive rate) とは、仮説検定において帰無仮説 (差なし) が正しいのに、対立仮説 (差あり) と判断してしまう割合である (第一種の過誤確率でもある)。一般に、仮説検定は偽陽性のみをコントロールしている。有意水準を 0.05 と

して  $p$  値  $< 0.05$  で棄却判断することは、偽陽性割合 = 0.05 に制御している。すなわち、有意水準 = 偽陽性割合である。間違える確率 (偽陽性割合) は 5% 残っている (0% が良いというわけではない)。T 値も  $p$  値も偽陽性割合を制御するように閾値が設定される。0.05 というのは伝統的なだけである。また偽陰性は放置しているので  $P$  値  $\geq 0.05$  でも帰無仮説採択と言えない (偽陰性が高くなる可能性がある)。

Eklund ら<sup>4)</sup>では安静時だけを対象にしたので、帰無仮説 (task に対して効果なし) が真という設定となり、効果ありと出てしまった解析結果を偽陽性と呼んだ。症例の無作為抽出法により 1000 回の解析を行い、そのうちに検定の結果として脳全体のどこかで帰無仮説を棄却した (偽陽性が起こった) 回数を 1000 で割ることによって偽陽性割合を計算した。この偽陽性割合が Family Wise Error である。SPM, FSL, AFNI といったソフトウェアによる解析結果を評価している。また本論ではクラスター推測に焦点を当てるが、Eklund ら<sup>4)</sup>ではボクセルレベルの推測も評価しており、偽陽性割合は宣言された 0.05 を満たしていた。

“ $P=0.001+10$  voxels” 法というのが、データ解析においてよく使われていたようである。これは CDT を 0.001 に設定して、サイズ閾値を 10 にするものである。この 10 という閾値は有意水準を 0.05 にしているわけでもなく、いきつくところとして、第 1 種の過誤 (偽陽性) 割合も 0.05 に保証しているわけでもない。このために場合によっては偽陽性割合が 70% にもなり得るのである (Eklund ら<sup>4)</sup>の Fig. 2)。式(1)に 3 節の設定で  $s=10$  を代入したところ、 $p=0.5857$  と算出された。偽陽性割合が 59% になることを表している。こうして有意水準を設定した上でクラスターサイズの閾値を設定すべきである。

Eklund ら<sup>4)</sup>の出版後に様々な対応がされた。文献<sup>8)</sup>では次の点を主張している；(1) a high cluster forming threshold (=CDT) を適切に設定すべき、(2) smoother than the size of voxels を適切にすべき、(3) One-sample test を実験するのは無意味。(1), (2)は確率場理論の仮定どおりであり、Eklund ら<sup>4)</sup>では数値的に示

してくれたこととなると述べられている。Organization for Human Brain Mapping (OHBM) の反応としてブログ ([http://www.ohbmbrian-mappingblog.com / blog / keep-calm-and-scan-on](http://www.ohbmbrian-mappingblog.com/blog/keep-calm-and-scan-on)) に見解を公表している。まとめとしては、(a) CDT をいくら高く設定しても、10 ボクセルで閾値をするのは良くない(b) permutation test は FWE を制御できる(c)どのソフトウェアにおいても CDT=0.01 にするのは良くない(d) CDT = 0.001 と設定しても場合によっては FWE を制御できない場合もある。文献<sup>8)</sup>とほとんど同じであるが、(d)については文献<sup>8)</sup>の(3)の主張を支持する。さらには Nichols が 40,000 の fMRI 研究の結果が無意味であるという Eklund ら<sup>4)</sup>の内容を、3,500 の fMRI 研究であると修正していることも述べられている。そして文献<sup>9)</sup>では解析の設定などに関してチェックリストとして使用できそうな表をまとめている。CDT の設定に関しては文献<sup>10)</sup>においても既に重要性が指摘されており、部分クラスターも結果として表示することを推奨している。解析ソフトウェアの一つである AFNI にはバグがあったがそれは既に修正しており、Eklund ら<sup>4)</sup>ではそのバグに関しても取り上げられていたために、AFNI group も見解を示した上で新しい方法論を発展させている<sup>11)</sup>。

これらから、そして前節までの内容から、まとめとして、CDT を高めに設定すること、平滑化は行う必要があること、クラスターサイズ検定においては有意水準に対応する閾値を使用することを注意しておく方が良いと考えられる。Eklund ら<sup>4)</sup>の結果からは平滑化に関しては FWHM を 8 mm で行っておけば問題なさそうであるが、CDT に関してはより注意が必要である。本論での解析例では、functional と structural MRI の違いがあるかもしれないが、CDT = 3.3 ( $p = 0.001$ ) よりむしろ CDT = 4.5 や 5 くらいまで高くした方が良いようである。高すぎて困ることは検出力が落ちることだと考える。permutation test が優良であるという結論もある。これはパラメトリック検定とノンパラメトリック検定の違いにもなり、一長一短あるわけであり完全にどちらが良いという結論には至らないと考える。また文献<sup>8)</sup>にもあ

るように one-sample の結果に関しては解釈が容易でないと考える。

### 関連する他の話題

本論で解説した SPM で伝統的に使用されていたクラスターサイズの検定方法は、平滑化された画像からの T-map に対し CDT を与えた上で行う検定方法であり、その選択に結果が大きく影響を受けることを見てきた。これに対し、閾値に依存しない方法が提案されている。文献<sup>12)</sup>は threshold-free cluster enhancement (TFCE) 法を提案している。文献<sup>13)</sup>では比較研究を行い TFCE 法の有用性を示している。文献<sup>14)</sup>は connectivity-based fixel enhancement (CFE) として TBSS の枠組みに拡張している。文献<sup>15)</sup>では皮質解析に拡張している。TFCE 法は FSL や CAT12 において実行することができる。

平滑化の方法もガウス関数に基づくが、どの領域に対しても同一のガウス関数を適用していることは限定的かもしれない。より柔軟な平滑化の方法も提案されている。ウェーブレットに基づく方法は WSPM toolbox<sup>16)</sup>によって実行することができる。

### ま と め

本論では脳画像解析におけるクラスター推測について、多くのユーザーがいると見込んだ SPM で採用されている方法でもある、クラスターを形成した上で有意なクラスターサイズを検出するための検定方法について解説を行った。クラスター推測は近年世界中で議論になっている。方法論の複雑さとソフトウェアの簡便さにより、作った方と使う方とで考え方が異なる部分もあるように感じた。その乖離を少しでも埋めることができるように p 値の計算式を比較的計算しやすい形に簡略化し、その挙動を実 VBM 解析の結果を通して示した。既存の議論と本論と共通することとしては、クラスター形成そして検定における閾値決めにも必要となる CDT が結果に大きく影響することである。クラスターを形成する際に、FWHM はソフト

のデフォルトである 8 mm などにして、視覚的にはなるが、ある程度大きくなりすぎないほどのクラスターができるように CDT を設定する必要がある。あとはそれに対応するサイズ閾値を、有意水準を設定した上で計算して、それ以上のサイズのクラスターを結果として報告すれば良い。

一般に解析は検証的と探索的に大分される。全脳の脳画像解析はあくまでも探索的な解析であると考えている。この場合は検出力がないと次の段階に進めないのが、厳密すぎずある程度で第 1 種の過誤が制御できていれば良いと考える。偽陽性を実験や文献調査など何らか別の方法で検証した上で進むことが良いと考える。より強固な証拠とするためには、メタアナリシスが有効な方法になると考えられる。そのためには解析結果の詳細を報告すべきである。また脳画像解析に限ったことではないが論文などで解析結果を公表する際には、limitation (特にサンプルサイズ) に言及した上で、過剰すぎない結論にすべきである。統計解析はソフトウェアの出現により簡易に実行できるようになっている反面に Eklund ら<sup>4)</sup>で取り上げられたような落とし穴も存在するかもしれない。方法をよく理解する、専門家と相談する、などが確実な方法かもしれないが、どのように解析したかを明示して再現できるようにすることも重要である。これによって研究条件が吟味され今後の研究に活かされる可能性もあり、無意味な研究結果とはならないだろう。

## 付 録

$T$  を検定統計量とする。  $S$  をクラスターサイズ (クラスター内のボクセル数) とする。この  $S$  に関する検定を行うために、まず  $S$  の分布関数が必要となる。次が成り立つ<sup>17)</sup>。

$$S^{2/D} \sim \text{Exp} \left( \left[ \frac{E(S)}{\Gamma(D/2+1)} \right]^{2/D} \right)$$

ただし、  $\text{Exp}(\lambda)$  は平均  $1/\lambda$  の指数分布を表す。  $E(S)$  は  $S$  の期待値、  $\Gamma(\cdot)$  はガンマ関数を表す。  $D$  は確率場の次元を表すものであ

り、ここで全脳解析を想定しているので 3 次元確率場を考えていることになる ( $D=3$ )。ここから  $p$  値を計算するには帰無仮説の下での  $E(S)$  を求め、それを代入した上で  $\Pr(S>s)$  をこの指数分布に基づき計算すればよい。ここでの  $s$  は観測されたクラスターサイズでありこれに対応する  $p$  値が求まることになる。また参考として  $X \sim \text{Exp}(\lambda)$  のとき  $\Pr(X>x) = \exp(-\lambda x)$  である。

はじめに帰無仮説の下での  $E(S)$  を求める。  $N$  を閾値を超えたボクセルの総数とする。  $L$  をクラスター数とし、解析対象領域 (ほとんどの場合が全脳) における検定統計量は滑らかであると仮定すると、クラスターサイズとクラスター数は独立である。こうして帰無仮説の下での  $E(S)$  は次のように与えられる。

$$E(S) = \frac{E(N)}{E(L)}$$

この式を詳しく見ていく。まずは分子  $E(N)$  から、

$$E(N) = V \Pr(T>h)$$

ただし、  $V$  は解析対象領域サイズを表し、  $h$  は CDT を表す。  $\Pr(T>h)$  は  $T$  に対する裾確率を表すので、  $T$  の分布として標準正規分布 ( $t$ -分布の近似) を仮定すると  $\Pr(T>h) = 1 - \Phi(h)$  と書き表すことができる。ただし  $\Phi(h)$  は標準正規分布の分布関数である。この式はさらに標準正規分布の密度関数  $\phi(h)$  によって  $1 - \Phi(h) \approx \phi(h)/h$  と近似できる。

次に分母  $E(L)$  についてであるが、これはオイラー標数の期待値によって近似される。オイラー標数はある閾値  $h$  を越える確率場であり、ここでは  $\chi_h$  と表す。その期待値は一般形が明示的に与えられる<sup>18)</sup>。

$$E(\chi_h) = \sum_{d=0}^D R_d \rho_d(h)$$

ただし、  $R_d$  は  $d$  次元における RESEL (RESolution ELEMENT) カウントであり、  $\rho_d(h)$  はオイラー標数密度である。  $E(L)$  は  $E(L) \approx R_3 \rho_3(h)$  と近似される。ここで、

RESEL は統計量の変動を表し、 $r$  と表すと次のように定義される。

$$r = \text{FWHM}_x \times \text{FWHM}_y \times \text{FWHM}_z$$

ただし、FWHM は各方向の Full Width at Half Maximum であり、SPM においては GLM の残差から推定される。 $R_3 = V/r$  と定義され、

$$\frac{R_3}{V} = \frac{1}{r} = \frac{1}{\text{FWHM}_x \times \text{FWHM}_y \times \text{FWHM}_z}$$

という関係式が成り立つ。標準正規分布を仮定すると  $E(L) \approx R_3 \times (4 \log 2)^{3/2} \times (h^2 - 1) \exp(-h^2/2) (2\pi)^{-2}$  となる。この式を標準正規分布の密度関数  $\phi(h) = \exp(-h^2/2) (2\pi)^{-1/2}$  を用いて、 $R_3 \times (4 \log 2)^{3/2} \times (h^2 - 1) \exp(-h^2/2) (2\pi)^{-2} = R_3 \times (4 \log 2)^{3/2} \times (h^2 - 1) \phi(h) (2\pi)^{-3/2}$  と書き直すことができる。さらにここで  $a = (4 \log 2/2\pi)^{3/2}$  とする。

以上のことから

$$\begin{aligned} E(S) &= \frac{E(N)}{E(L)} \approx \frac{V\phi(h)/h}{R_3 a (h^2 - 1) \phi(h)} \\ &= \frac{1}{a} \times r \times \frac{1}{h(h^2 - 1)} \end{aligned}$$

と書くことができる。そして冒頭の指数分布に反映させる。

$$\begin{aligned} p \text{ 値} &= P(S > s) = P(S^{2/3} > s^{2/3}) \\ &= \exp\left(-\left[\frac{E(S)}{\Gamma(3/2 + 1)}\right]^{-2/3} s^{2/3}\right) \\ &= \exp\left(-\left[\frac{\frac{1}{a} \times r \times \frac{1}{h(h^2 - 1)}}{\Gamma(5/2)}\right]^{-2/3} s^{2/3}\right) \\ &= \exp\left(-b \left[\frac{s}{r} h(h^2 - 1)\right]^{2/3}\right) \end{aligned}$$

ただし  $b = (a\Gamma(5/2))^{2/3} = (4 \log 2 / 2\pi) \Gamma(5/2)^{2/3}$  は定数であり、 $b = 0.5334942$  である。こうして式(1)が得られ、 $p$  値は CDT  $h$  と確率場の滑らかさに関する  $r$  の関数で与えられ

ることが分かる。またこの  $p$  値の式を  $s$  について解くと

$$\begin{aligned} -\frac{1}{b} \times \log p \text{ 値} &= \left[\frac{s}{r} h(h^2 - 1)\right]^{2/3} \\ \Rightarrow s &= \frac{r[-\log p \text{ 値}/b]^{-2/3}}{h(h^2 - 1)} \end{aligned}$$

となり式(2)を得る。こうしてこの式の  $p$  値に有意水準を代入するとそれに対応するクラスターサイズの閾値を計算することができる。

## 文 献

- 1) 根本清貴. すぐできるVBM精神・神経疾患の脳画像解析 SPM12対応DVD付. 学研メディカル秀潤社, 2014
- 2) 川口 淳: 脳MRIデータの統計解析, 計量生物学 2013; 33: 145-174
- 3) 川口 淳, 角間辰之: 多重比較とは何か. Diabetes Frontier 2014; 25: 733-741
- 4) Eklund A, Nichols TE, Knutsson H: Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci USA 2016 113: 7900-7905
- 5) Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J: Unified univariate and multivariate random field theory. NeuroImage 2004; 23: S189-S195
- 6) Ashby FG. Statistical Analysis of FMRI Data., MIT Press, 2011
- 7) Chung MK. Statistical and Computational Methods in Brain Image Analysis. CRC press, 2013
- 8) Flandin G, Friston KJ. (2016). Analysis of family-wise error rates in statistical parametric mapping using random field theory. arXiv preprint arXiv: 1606.08199
- 9) Nichols TE, Das S, Eickhoff SB, et al. Best practices in data analysis and sharing in neuroimaging using MRI. Nat Neurosci 2017; 20: 299-303
- 10) Woo CW, Krishnan A, Wager TD: Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. NeuroImage 2014; 91: 412-419
- 11) Cox R, Chen G, Glen DR, Reynolds RC, Taylor PA: FMRI Clustering in AFNI: False Positive Rates Redux. Brain Connect 2017; 7: 152-171
- 12) Smith SM, Nichols TE: Threshold-free cluster

- enhancement : addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* 2009 ; 44 : 83–98
- 13) Li H., Nickerson LD, Nichols TE, Gao JH : Comparison of a non-stationary voxelation-corrected cluster-size test with TFCE for group-Level MRI inference. *Hum Brain Mapp* 2017 ; 38 : 1269–1280
  - 14) Raffelt DA, Smith RE, Ridgway GR, Tournier JD, Vaughan DN, Rose S, Henderson R, Connelly A : Connectivity-based fixel enhancement : whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres. *NeuroImage* 2015 ; 117 : 40–55
  - 15) Lett TA, Waller L, Tost H, et al. : Cortical surface-based threshold-free cluster enhancement and cortexwise mediation. *Hum Brain Mapp* 2017 in press
  - 16) Van De Ville D, Seghier ML, Lazeyras F, Blu T, Unser M : WSPM : wavelet-based statistical parametric mapping. *NeuroImage* 2007 ; 37 : 1205–1217
  - 17) Nosko VP. 1969. Local structure of gaussian random fields in the vicinity of high level shins. *Sov Math : Doklady* 10, 1481–1484
  - 18) Worsley KJ, Marrett S, Neelin P, Evans AC : Searching scale space for activation in PET images. *Hum Brain Mapp* 1996 ; 4 : 74–90

## Statistical Cluster Inference in Brain Imaging Analysis

Atsushi KAWAGUCHI

*Center for Comprehensive Community Medicine  
Faculty of Medicine, Saga University  
5-1-1 Nabeshima, Saga 849-8501*

Statistical cluster inference is integral to brain imaging analysis. We briefly review the most commonly used method for brain imaging analysis, namely, the cluster size test, which is based on the random field theory. Statistical software packages, such as Statistical Parametric Mapping (SPM), allow for easy implementation of such an analysis ; however, the user is cautioned to use the software application in the appropriate setting. This article provides the p-value formulation with a simple form, an explanation of how the software worked, and the appropriate settings for the computation with applications to real data. We also provide a review about recent discussions and extensions for the cluster inference in brain imaging analysis.