

多重比較補正

京都大学大学院医学研究科

医学統計生物情報学

川口 淳

- ① 統計学的仮説検定
- ② 多重検定とその補正
- ③ 画像解析における多重補正法
- ④ 付録

Random Field Theory

包括脳・MRI脳画像解析チュートリアル
2014年12月13日

1

統計学的仮説検定

2

2つの仮説と検定統計量

- 仮説検定を行うには、2つの仮説と検定統計量が必要！
- 2つの仮説 (Hypothesis)
 - » 【帰無仮説 (Null Hypothesis) H_0 】
 - > 多くの場合は差 (効果) が無い事を表す
 - » 【対立仮説 (Alternative Hypothesis) H_A 】
 - > 多くの場合は導きたい結論 (差がある) を表す
 - » H_0 を棄却 (reject) する (H_A を採択) か否かを検定する
 - > H_0 に着目するので、検定を表すときは H_0 のみを記述する事が多い
- 検定統計量 (Test Statistic) T
 - » H_0 の根拠を集約していると見なせる。
 - » 多くの場合、 H_0 が真であれば T は小さく、偽であれば大きい。

3

単回帰モデルの場合

- 単回帰モデル： $y = \beta_0 + \beta_1 x + \varepsilon$
- 調べたいこと： y と x に関係があるか？

$$\text{仮説：} \begin{cases} \text{帰無仮説：} \beta_1 = 0 & \Rightarrow y \text{と} x \text{に関係がない} \\ \text{対立仮説：} \beta_1 \neq 0 & \Rightarrow y \text{と} x \text{に(直線)関係がある} \end{cases}$$

- 検定統計量 (T 値)

$$T = \frac{\hat{\beta}_1 - 0}{\text{SE}(\hat{\beta}_1)}$$

帰無仮説の下では、 $\beta_1 = 0$

- » $\text{SE}(\hat{\beta}_1)$ は $\hat{\beta}_1$ の標準誤差。実際にはデータから算出するための公式アリ。
- どのようにして帰無仮説の棄却を判断する？
 - » T が大きいと帰無仮説が棄却できる。

4

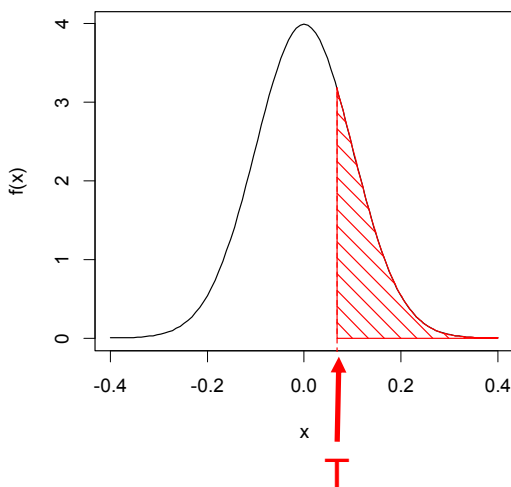
検定統計量の分布

- 統計学的な（確率に基づく）判断をするために帰無仮説の下でのTの分布が必要
- 帰無仮説が正しいと仮定して分布を決める
 - » 帰無分布とも呼ばれる
 - » 中心極限定理から標準正規分布を用いることが多い
 - » あとは、正規分布から派生するt分布やF分布

5

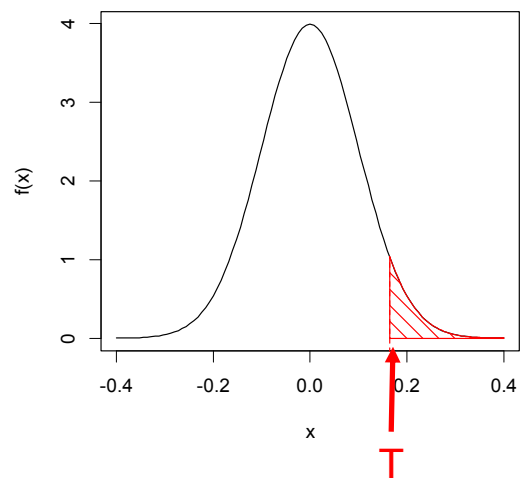
帰無分布による判断

Tが小さい



分布下の曲線下面積(= 裾確率)が大きい

Tが大きい



分布下の曲線下面積(= 裾確率)が小さい

6

p値と有意水準

➤ **p値** = $\Pr(T > t | H_0)$

» H_0 が真であると仮定して、データから計算された
Tの値 t を確率的に評価（小さい or 大きい）

➤ **有意水準** $\alpha = \Pr(T > u_\alpha | H_0)$

» どれくらいp値が小さい（ T が大きい）と棄却で
きるか

» u_α ：確率が α になるような閾値

7

帰無仮説の棄却判断

➤ 有意水準 $\alpha = 0.05$ （5%）での判断

» $p\text{値} < 0.05 (= \alpha)$ or $t > u_\alpha$

> $\Rightarrow$ 有意水準5%で H_0 が棄却され、 H_A が成り立つ(p値=xxx)

» $p\text{値} \geq 0.05 (= \alpha)$ or $t \leq u_\alpha$

> $\Rightarrow$ 有意水準5%では、 H_0 か H_A かについては、何もいえない

➤ T値とp値は1対1の関係だったので、どちらでも判断できる。

» $p\text{値} = 0.012 < 0.05 \Leftrightarrow T\text{値} = 2.5 > 1.96$

8

2種類の過誤

	真	
	帰無仮説が真	対立仮説が真
検定結果	帰無仮説を採択 ($p\text{値} \geq 0.05$)	正しい 第2種の過誤 false negative
	対立仮説を採択 ($p\text{値} < 0.05$)	第1種の過誤 false positive 正しい

- 第1 (2) 種の過誤確率を α (β) とする
- α エラー, β エラーとも呼ばれる

9

過誤確率の制御

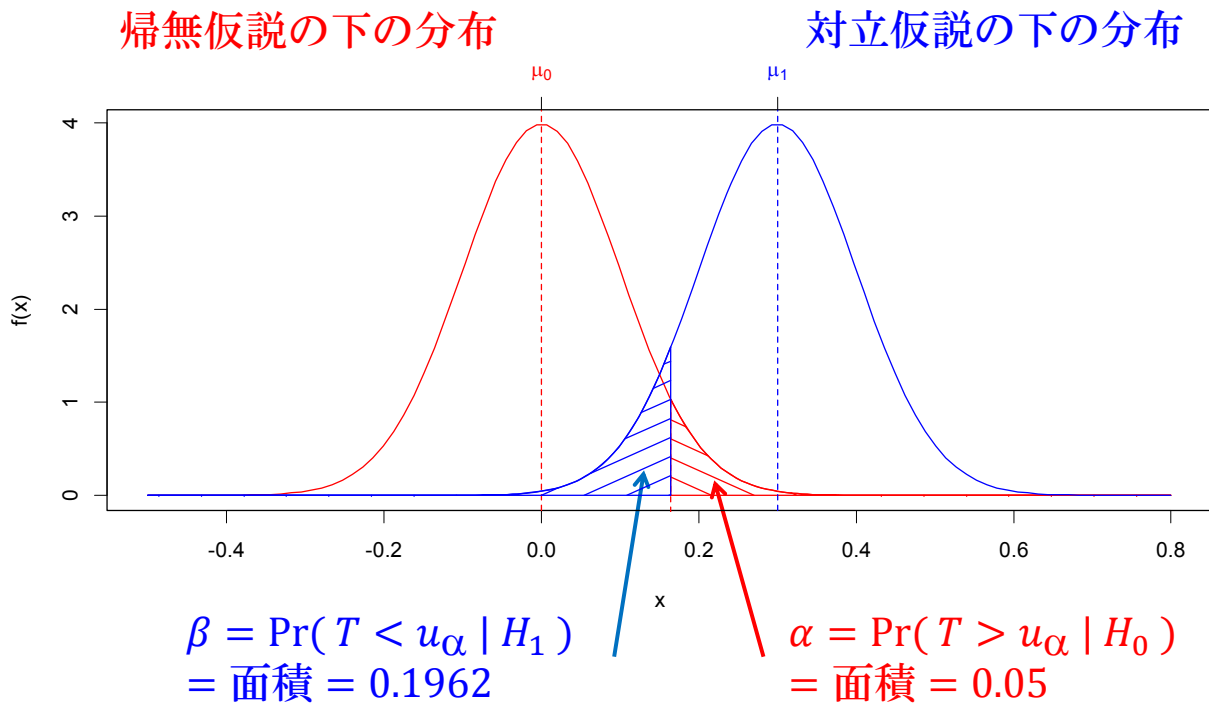
- 仮説検定は α のみをコントロールしている (β は放置)
 - » $P\text{値} \geq 0.05$ でも帰無仮説採択と言えない (β が高くなる可能性がある)
- $p\text{値} < 0.05$ で棄却判断する事は, $\alpha=0.05$ に制御している

有意水準 = 第1種の過誤確率

- » 間違える確率(false positive)は5%残っている
- $p\text{値の閾値}\alpha$ で, $T\text{の閾値}u_\alpha$ で, } 第1種の過誤確率を α に制御している
- 過誤確率が0.05というのは伝統的なだけ。

10

過誤確率の図示



11

多重検定とその補正

12

多重性について

- ボクセル毎に検定が行われるとき，
 - » ボクセル数分だけ検定が行われ，有意なボクセルたちを検出する
- どの100,000個のボクセルが有意？
 - » $\alpha=0.05 \Rightarrow 5,000$ 個のボクセルがfalse positive
- どの100個のクラスターが有意？
 - » $\alpha=0.05 \Rightarrow 5$ のクラスターがfalse positive

13

族 (family)

- 多重比較で重要な事として，どのような仮説の集合において過誤の計算に考えるか，がある。
 - » 帰無仮説が複数あると考えてよい。結論は集合単位。
- この仮説の集合を，族 (family) と呼ぶ
 - » 例えば，多群(の組み合わせ)，測定時期，評価項目の水準，ボクセル集合，など
- 族単位の過誤確率をFamily Wise Errorと呼ぶ
- 族の設定は主観的ではあるが，その設定によって結果が異なる。

14

過誤 (Error) の種類

- 何か閾値を設け、 m 個の仮説検定を行った場合、下のような分割表を作成できる。

	帰無仮説を採択	帰無仮説を棄却	計
帰無仮説が真	U	V	m_0
帰無仮説が偽	T	S	m_1
計	W	R	m

- ① Familywise Error Rate (FWER)
 - » Familywise Error(FWE)は、複数の検定結果のうち1つ以上偽陽性がある事
 - » FWER はその確率

$$\text{FWER} = \Pr(\text{FWE}) = \Pr(V \geq 1)$$
- ② False Discovery Rate (FDR)
 - » 棄却された帰無仮説のうちの偽陽性率

$$\text{FDR} = E \left[\frac{V}{R} \right]$$
 - » R 個のボクセルを有意とした場合、 V 個が誤り
 - » E は期待値を表す

15

過誤を評価するには

- p 値または T 値の閾値を変えれば良い。
- どのように？
 - » それぞれの過誤確率が水準を満たすように
- そういうわけで、多重補正とは
 - » 族の構成員で計算された p 値や T 値たちの、閾値を与え直すことである。
 - » 多くの補正法が存在するが、閾値が異なるだけ、と考えて良い
- 逆に、閾値をそのままにするのであれば、 p 値や T 値たちの値の方を「補正」する。

16

FWERの計算

帰無仮説を2つとする

帰無仮説	検定統計量	棄却条件	有意水準	第1種の過誤の確率
H_{01}	T_1	$T_1 > c$	0.05	$\Pr(T_1 > c H_{01}) = 0.05$
H_{02}	T_2	$T_2 > c$	0.05	$\Pr(T_2 > c H_{02}) = 0.05$

- T_1 と T_2 は独立として, FWERを計算

$$\begin{aligned}\Pr(T_1 > c \text{ or } T_2 > c) &= 1 - \Pr(T_1 \leq c \text{ and } T_2 \leq c) \\ &= 1 - \Pr(T_1 \leq c) \Pr(T_2 \leq c) \\ &= 1 - (1 - 0.05) \times (1 - 0.05) = \mathbf{0.0975}\end{aligned}$$

- FWERは0.0975になり, 0.05を上回っている
 - » これは, 有意水準0.05の検定を繰り返したためである
 - » type 1 FWERを0.05になるように補正する

17

ボンフェロニーの補正

- 最も単純なFWER補正法
 - » p値は, 検定の数かける
 - » 有意水準は, 検定の数で割る
- 先ほどの例で, 各々の有意水準0.05を1/2倍(=0.025)する
 - » 水準を厳しく(帰無仮説を棄却し難く)
$$\Pr(T_1 > c \text{ or } T_2 > c) = 1 - 0.975 \times 0.975 = \mathbf{0.049375}$$
 - » Type 1 FWERを0.05に制御できた
- **独立性を仮定している**ので, 保守的になりやすい
 - » 保守的=パワー不足

18

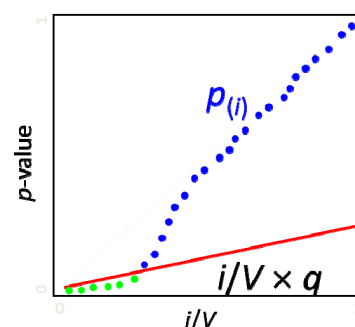
FDRの制御方法

➤ Benjamini & Hochberg (BH)法

- » 閾値 q （通常0.05）が所与
- » p 値を小さい順に並び替える, $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$
- » 次を満たす最大の i を r とする

$$p_{(i)} < i/m \times q$$

- » $p_{(1)}, p_{(2)}, \dots, p_{(r)}$ に対応する帰無仮説を棄却する



傾き q , 切片0の直線の下にある p 値は有意であると判断する

➤ BH法における棄却は, FDRを q 以下に制御する

$$\text{FDR} = q \frac{m_0}{m} \leq q$$

➤ $p_{(i)}$ に対応する p -value (FDR補正済みの p 値)

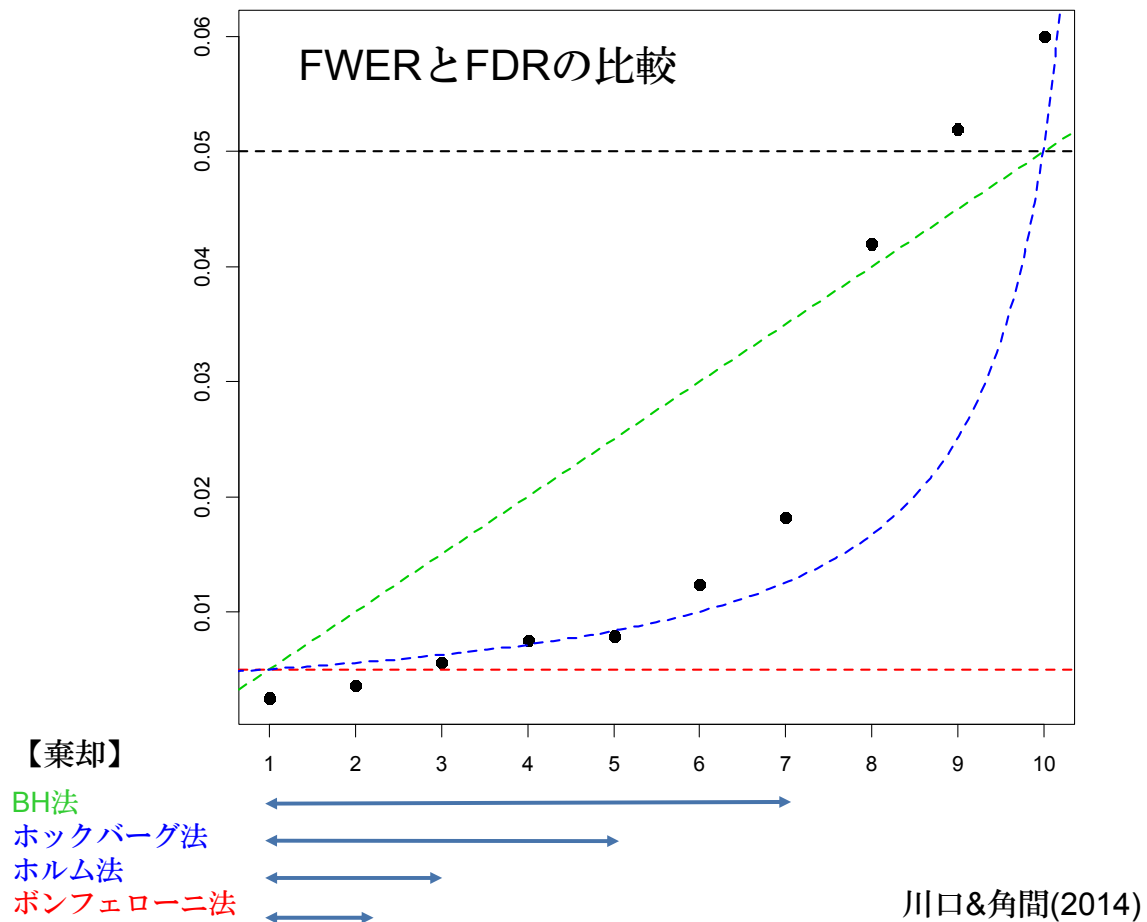
$$\tilde{p}_{(i)} = \min_{j \geq i} (m p_{(i)} / j)$$

19

FWERとFDRの比較

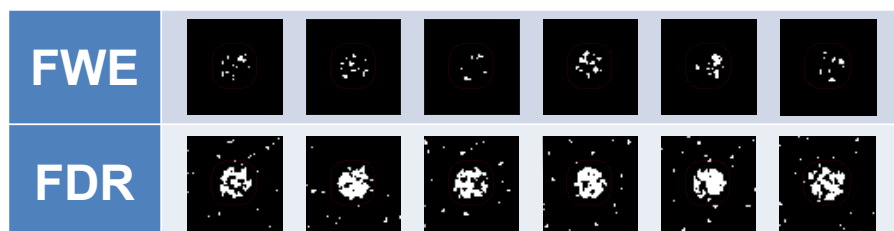
10個の帰無仮説の p 値に対する各方法による補正済有意水準（閾値）と検定結果

順位	元の p 値	ボンフェローニ (FWER 0.05)		BH (FDR 0.05)	
		閾値	帰無仮説	閾値	帰無仮説
1	0.0025	$1/10 \times 0.05 = 0.0050$	棄却	$1/10 \times 0.05 = 0.0050$	棄却
2	0.0036	$1/10 \times 0.05 = 0.0050$	棄却	$2/10 \times 0.05 = 0.0100$	棄却
3	0.0056	0.0050		0.0150	棄却
4	0.0075	0.0050		0.0200	棄却
5	0.0079	0.0050		0.0250	棄却
6	0.0124	0.0050		0.0300	棄却
7	0.0182	0.0050		0.0350	棄却
8	0.0420	0.0050		0.0400	
9	0.0520	0.0050		0.0450	
10	0.0600	$1/10 \times 0.05 = 0.0050$		$10/10 \times 0.05 = 0.0500$	



21

FWERとFDRの比較のまとめ



- FDRの方がFWERに比べより多くの有意なボクセルを検出しがちである
 - » 検定数が少ないときは、そうとは言えない場合がある
 - » $V = R$ のときは、 $FDR = FWER$
- 実践的にはFWERは保守的になりすぎることから閾値決めとしてはFDRの方が有用である。

22

画像解析における多重補正法

23

FWERと検定統計量の最大値

- FWERを制御する多重補正については,

$$\text{FWER} = \Pr(\text{FWE}) \text{ "FWERはFWEのおこる確率"}$$

$$= \Pr(\text{"1つ以上の偽陽性がある"}) \text{ "定義より"}$$

$$= \Pr(T \text{の最大値} > c | H_0) \text{ "帰無仮説が正しいのに棄却"}$$

- 帰無仮説が正しいという状況の下, 検定統計量の最大値が閾値を超えると, FWEが起こった事になる.

$$\gg \text{棄却} = T > c \Rightarrow T \text{の最大値} > c \Rightarrow \text{少なくとも一つが棄却}$$

- T (たちの)の分布ではなく, T の最大値の分布を考えれば良い.

24

最大値による制御

- FWERは最大値の分布として表される

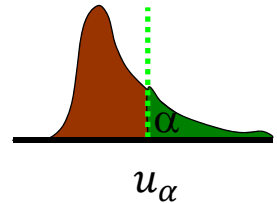
$$\text{FWER} = \Pr(\text{FWE}) = \Pr\left(\bigcup_i \{T_i \geq u\} \mid H_0\right) = \Pr\left(\max_i \{T_i\} \geq u \mid H_0\right)$$

- 最大値の分布の $100(1 - \alpha)\%$ 点を閾値にする ($u = u_\alpha$)

» FWERを $100\alpha\%$ に抑えることができる

$$\text{FWER} = \Pr\left(\max_i \{T_i\} \geq u_\alpha \mid H_0\right) = \alpha$$

$$\text{ただし, } u_\alpha = F_{\max}^{-1}(1 - \alpha)$$



- 最大値の分布に対する閾値 u_α は、Random Field Theory (付録) によって求められる。

» ボクセル間の相関を考慮している

25

Resampling

- 分布を仮定しない多重補正の方法
- ラベルをランダムにシャッフルして、検定を行う。これを繰り返し行い得られた検定統計量の分位点を閾値に用いる
- 計算機により実行される
- Winkler et al. (2014 Neuroimage)

26

なるべく検定を減らす

- 検定数を減らせば、補正も厳しくなくて良い
- ボクセルレベル
 - » 【長所】 仮定が少ない自然な解析
 - » 【短所】 全ボクセルが対象なので、補正が厳しい
- クラスターレベル
 - » 【長所】 検定数は減り、補正を緩やかにできる（power up）
 - » 【短所】 クラスター決めがやや主観的（height決め）
- Small Volume Correction
 - » 【長所】 検定数は減り、補正を緩やかにできる（power up）
 - » 【短所】 （説得力のある）事前知識が必要

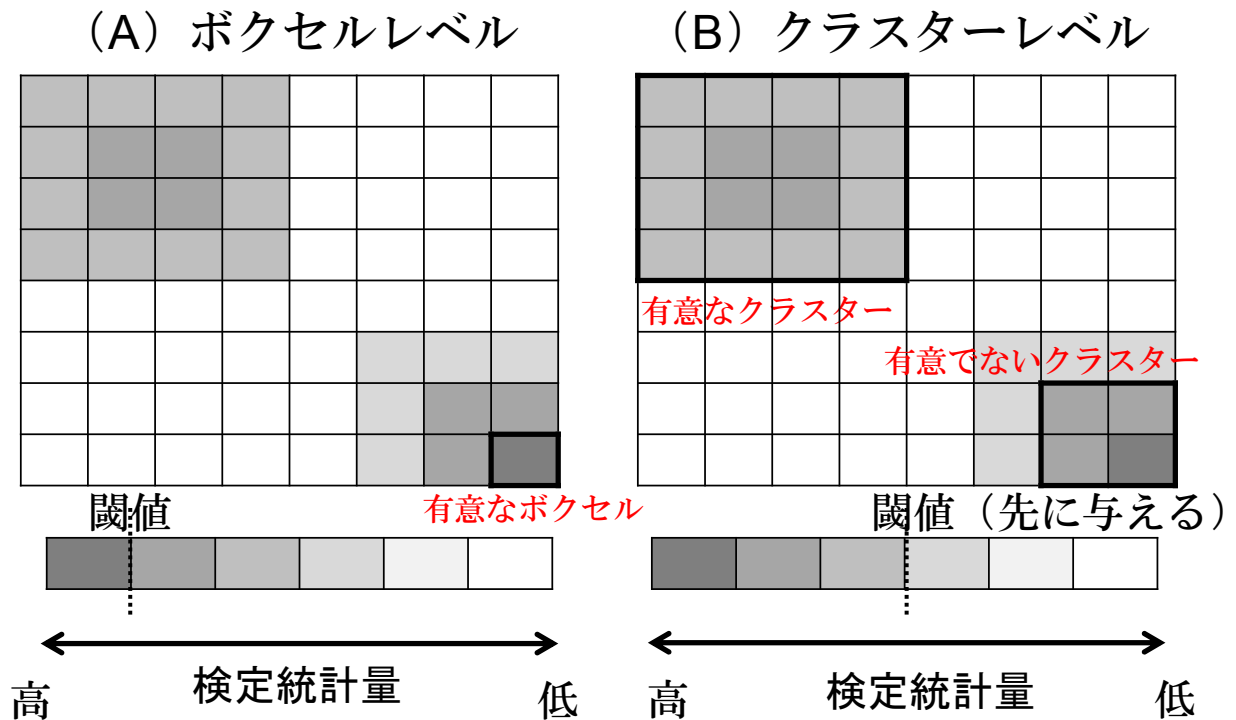
27

クラスター

- ボクセルレベルの推測だと．．．
 - » 多くのボクセルを考える事になり、多重補正においても高めの閾値が設定される（検出されにくい）
 - » 近隣のボクセルは似たような結果になるのが自然だが、ボクセルレベルだと近隣情報は無視される
- なので、ボクセルの集まり（クラスター）を考える

28

クラスターレベル



29

Cluster-level法の概要

- ① SPM (z-map) を作成.
- ② zに対する閾値 (height threshold) を決める
 - » ただし高め ($T=2.5$ とか 3)
- ③ その閾値を超えたボクセルでクラスターを形成
 - » 隣接するボクセルの塊を一つのクラスターと考える
 - » SPMでは18-connectivity を使用
- ④ どのクラスターが有意かをp値に基づき判断する
 - » 帰無仮説は、すべてのボクセルで差なし
 - » クラスターが大きければ、帰無仮説棄却.
 - » 有意なクラスター内では、容積に有意差あり

30

まとめ

- 多重比較補正法は，さまざまある。
 - » FWER, FDR, cluster level,...
- 検定のを減らし，ある程度の検出力を持ったもので良い
 - » VBMなどは探索的解析だと考える。
 - > のちに検証的な解析，実験があるとして。
- このような観点から，cluster levelのFDRが有用ではないかと考える。

31

参考文献

- Poldrack R. A., Mumford J. A., Nichols T. E. (2011). Handbook of Functional MRI Data Analysis. Cambridge University Press.
- 川口淳（2012）．脳MRIデータの統計解析．計量生物学会誌,33(2),145-174
- 川口淳，角間辰之（2014）．多重比較とは何か．Diabetes Frontier, Vol.25 No.6，印刷中
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, 92, 381-397.

32

確率場（Random Field）？

- 順序づけされた確率変数の集まり
 - » 順序づけは，脳画像ではボクセルの位置と考えられる。
- Gaussian Random Field (GRF)
 - » その確率変数の集まりの同時分布が多変量正規分布
 - » 相関構造を考えることができる
- 各ボクセルの検定統計量は，正規分布に従うと考えられる。
 - » T統計量であっても，p値をz-変換（SPMにおけるz-map）
- ここでは，3次元GRFを考える。

RFTによるFWE補正 (peak-level)

- ボクセルの検定統計量 t に対する補正済み p 値

$$p_{FWE-corr} = R \frac{(4 \log 2)^{3/2}}{(2\pi)^2} (t^2 - 1) \exp\left(-\frac{t^2}{2}\right)$$

» ただし, $R = (\text{RESEL Count}) = \frac{(\text{解析対象ボクセル数})}{(\text{RESEL=統計量の変動})}$

	$p_{FWE-corr}$	補正度合い	帰無仮説を
統計量 t の値が大きい	小	軽い	棄却し易い
対象ボクセル数 が多い	大	厳しい	棄却し難い
RESELが大きい	小	軽い	棄却し易い

35

RESEL

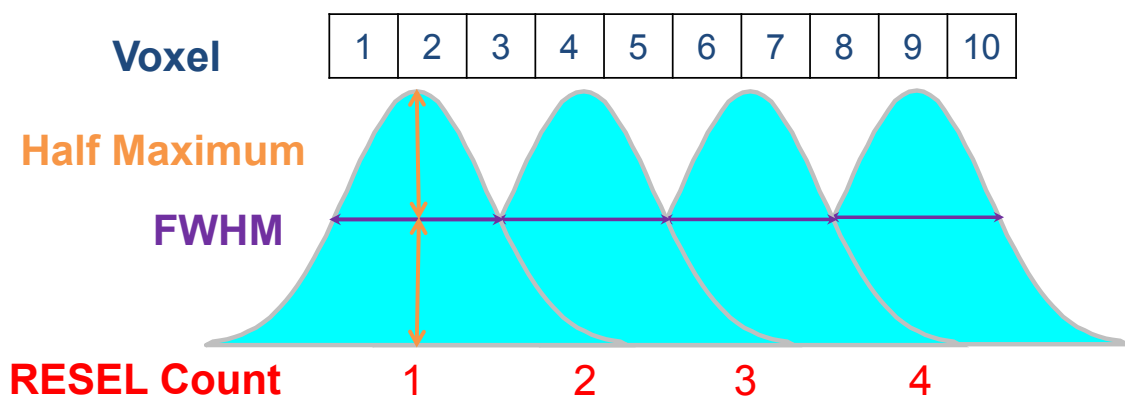
- **RESEL** (RESolution ELe ment) 統計量の変動を表す

$$1 \text{ RESEL} = \text{FWHM}_x \times \text{FWHM}_y \times \text{FWHM}_z$$

» FWHM = Full Width at Half Maximum

> GLMの残差 $y - X\hat{\beta}$ から推定される (spm_est_smoothness.m)

- $(\text{RESEL Count}) = (\text{解析対象ボクセル数}) / (\text{RESEL})$
- 【例】 10 voxels, 2.5 FWHM (=RESEL) $\Rightarrow$ 4 RESELS



36

RFTによるFWE補正 (cluster-level)

Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm		
<i>p</i>	<i>c</i>	<i>p</i> _{FWE-corr}	<i>q</i> _{FDR-corr}	<i>k_E</i>	<i>p</i> _{uncorr}	<i>p</i> _{FWE-corr}	<i>q</i> _{FDR-corr}	<i>T</i>	(<i>Z</i>)	<i>p</i> _{uncorr}			
0.000	22	0.000	0.000	1055	0.000	0.000	0.010	7.78	6.31	0.000	44	-28	-32
						0.001	0.083	6.73	5.68	0.000	50	-46	-26
						0.001	0.083	6.70	5.66	0.000	46	-54	-20

- cluster-level推測は**クラスターサイズ(k_E)**に基づく
- クラスターサイズに対する補正済みp値

$$p_{\text{FWE-corr}} = 1 - \exp(-E(\chi_u)p_{\text{uncorr}})$$

- » p_{uncorr} は補正無しのp値 ($q_{\text{FDR-corr}}$ の計算にも使われる)
- » $E(\chi_u)$ はオイラー標数であり, RESEL countと比例する

37

SPMでの出力

- RFTに基づく推測をする (p値を求める)

- » ① 解析対象部位の総容積 $\lambda(\Omega)$
- » ② 滑らかさ (元を辿ると, FWHM)

が必要である

Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm		
<i>p</i>	<i>c</i>	<i>p</i> _{FWE-corr}	<i>q</i> _{FDR-corr}	<i>k_E</i>	<i>p</i> _{uncorr}	<i>p</i> _{FWE-corr}	<i>q</i> _{FDR-corr}	<i>T</i>	(<i>Z</i>)	<i>p</i> _{uncorr}			
0.000	22	0.000	0.000	1055	0.000	0.000	0.010	7.78	6.31	0.000	44	-28	-32
						0.001	0.083	6.73	5.68	0.000	50	-46	-26
						0.001	0.083	6.70	5.66	0.000	46	-54	-20

Height threshold: $T = 5.75$, $p = 0.000$ (0.000)
 Expected number of clusters: $\lambda = 10.375$
 Expected number of clusters: $\lambda = 10.375$
 Expected number of clusters: $\lambda = 10.375$
 FWHM: 5.014, FWHM: 7.047, FWHM: 2.704, 131

Degrees of freedom = [1.0, 52.0]
 FWHM = 18.1 19.0 18.4 mm mm mm; 9.0 9.5 9.2 {voxels}
 Volume: 6255848 = 781981 voxels = 952.7 resels
 Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 792.86 voxels)

残差より推定

①

mask.img
より求まる

Degrees of freedom = [1.0, 52.0]

FWHM = 18.1 19.0 18.4 mm mm mm; 9.0 9.5 9.2 {voxels}

Volume: 6255848 = 781981 voxels = 952.7 resels

Voxel size: 2.0 2.0 2.0 mm mm mm; (resel = 792.86 voxels)

= Volume/Voxel size

= Volume/Voxel size

= 6255848/2³ ⇒ FDR-corr

⇒ FWE-corr

38