

S.P.M. for Dummies

Introduction au Random Field Theory

Jean-Etienne Poirrier*

9 janvier 2006

Table des matières

1	Introduction	3
1.1	Des cartes statistiques paramétriques à la 3D	3
1.2	Quelques définitions	4
1.3	Quels sont les problèmes?	4
2	Le problème de la comparaison multiple	5
2.1	L'hypothèse nulle en statistiques classiques	5
2.2	Application de l'hypothèse nulle en imagerie fonctionnelle	5
2.3	Les méthodes de test de l'hypothèse nulle liée à la famille	6
2.4	Seuil de hauteur et pouvoir de localisation	6
3	La correction de Bonferroni	7
3.1	Expression mathématique de la correction de Bonferroni	8
3.2	Exemple	9
3.3	La correction de Bonferroni n'est souvent pas applicable!	9
4	La corrélation spatiale	9
4.1	Corrélation due à la collecte des données	10
4.2	Corrélation due au pré-traitement	10
4.2.1	Ré-alignement, normalisation spatiale, ré-échantillonnage	10
4.2.2	Lissage	10
5	Bonferroni n'est pas applicable en imagerie fonctionnelle	10
5.1	Un peu de maths	11
5.2	Exemple	11
5.3	Le lissage et les observations indépendantes	13

*Ce texte est une version étendue, "brodée" en français autour des chapitres 13 et 14 du livre "Human Brain Function" de Ashburner, Friston et Penny [1]. J'ai essayé de traduire un maximum de termes techniques en français, même si parfois, cela frise le ridicule. Ce texte m'a permis de mieux préparer ma présentation au Centre de Recherches du Cyclotron de l'Université de Liège, le 11 janvier 2006. Je n'ai jamais touché ni à la résonance magnétique fonctionnelle ni à S.P.M. mais je pensais que cet "exercice de style" pourrait m'apprendre beaucoup sur ces techniques statistiques et, éventuellement, les appliquer à d'autres domaines d'analyse d'images en biologie. Bon, ce n'est pas pour tout de suite ... Sauf mention contraire (notamment les images du livre), tous les documents sont sous licence GNU FDL et disponibles à l'adresse <http://www.poirrier.be/jean-etienne/presentations/rft/>

6	La Random Field Theory	14
6.1	Le caractère lisse et les reseals	14
6.2	La caractéristique d'Euler	15
6.2.1	Exemple	15
6.2.2	Signification de la caractéristique d'Euler attendue	15
6.2.3	Calcul de la caractéristique d'Euler attendue	16
6.2.4	Application de la caractéristique d'Euler calculée à notre exemple	16
6.3	Les seuils trouvés par la RFT et la correction de Bonferroni	19
6.4	La RFT et l'imagerie fonctionnelle 3D	19
6.5	Les vraies dépendances de l'EC attendue	20
6.6	Les hypothèses régionales	22
7	Suppositions préalables	22
7.1	Cas où les suppositions préalables de la RFT ne sont pas respectées	22
8	La RFT, un pas plus loin	23
8.1	Le maximum de la statistique de test	23
8.1.1	Application de la correction de Bonferroni	23
8.1.2	Application de la RFT	24
8.1.3	Pratiquement	24
8.2	L'étendue spatiale maximale du test statistique	25
8.2.1	La recherche dans de petites régions	26
8.3	Quelques détails supplémentaires	26
8.3.1	L'estimation de la FWHM	26
8.4	La False Discovery Rate	27
9	Et où allons-nous, maintenant ?	27

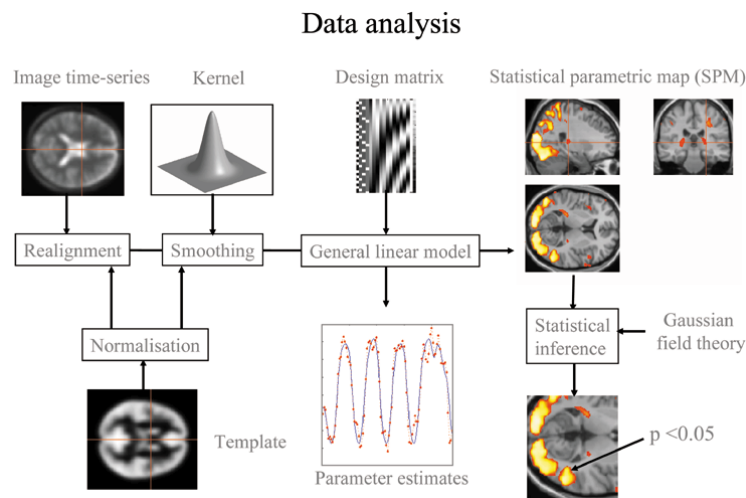


FIG. 1 – Schéma de processus de traitement en imagerie (image reprise du livre)

1 Introduction

Ce chapitre va traiter du problème des comparaisons multiples en imagerie fonctionnelle, de la manière de le résoudre, notamment grâce à la théorie du champ de hasard (*Random Field Theory* ou RFT).

Pour rappel, nous sommes dans l'imagerie en neurosciences. Ce domaine utilise des techniques comme la tomographie par émission de positron (*Positron Emission Tomography*, PET) ou l'imagerie fonctionnelle par résonance magnétique (*functional magnetic resonance imaging*, fMRI).

Dans les chapitres précédents, vous avez vu :

- la neuroanatomie “computationnelle”. Cette partie s’occupait du réaligement et de la normalisation spatiales. En effet, les données brutes de neuroimagerie contiennent des composantes de variances non désirées. Ces chapitres 1 à 6 vous ont permis de voir comment réduire ces composantes et les différentes manière de ré-aligner toutes les images variables acquises dans un espace anatomique standard. Les buts sont, bien sûr, de pouvoir aisément comparer ces données ainsi “standardisées” entre sujets, groupes, expériences, laboratoires, etc. mais surtout de préparer ces données à l’analyse proprement dite.
- la modélisation (de ce qui a été enregistré) est cette analyse, rendue possible surtout grâce au modèle linéaire général (*general linear model*, GLM). Cette modélisation est vue en détails aux chapitres 7 à 13.

Finalement, l’inférence statistique que nous allons voir (et qui suivra dans les chapitres suivants) permettra de passer d’un ensemble de cartes (indépendantes ?) à la détermination d’un effet dans l’ensemble de toutes ces cartes. Nous arrivons dans la partie en bas, à droite du processus de traitement illustré à la figure 1.

1.1 Des cartes statistiques paramétriques à la 3D

Comme je le comprends, les images obtenues sont généralement analysées sous forme de “cartes” de données statistiques. Ces cartes sont obtenues par des processus statistiques étendus à l’espace qui sont là pour tester les hypothèses d’effets spécifiques régionaux. On n’a pas de cartes de pixels comme le prendrait un appareil photo (donnant une image 2D dont chacun des éléments finis,

appelés pixels, représentent la projection d'une couleur d'un espace 3D en 2D). On a une carte (et même un volume) de voxels ¹.

Les cartes de statistiques paramétriques (*statistical parametric maps*, SPM) sont issues de ces processus statistiques et contiennent donc des voxels dont les valeurs sont distribuées selon une fonction de densité de probabilité connue (sous l'hypothèse nulle). Les distributions utilisées habituellement sont la T de Student et la F (respectivement “carte T” et “carte F”).

Dans une analyse d'imagerie fonctionnelle standard, nous faisons correspondre un modèle statistique aux données, afin de nous donner les paramètres du modèle. Nous utilisons alors ces paramètres pour chercher un effet qui pourrait nous intéresser (comme, par exemple, une différence entre une ligne de base et une tâche). Dans le cas de cette imagerie, pour chaque voxel du cerveau, nous calculons une valeur statistique qui est une sorte de test pour juger de l'intérêt de ce voxel. Au final, nous nous retrouvons avec un grand volume de valeurs statistiques.

Ainsi, énoncé encore autrement, chaque voxel est calculé à partir d'un test statistique standard univarié. Les paramètres statistiques résultants sont assemblés dans une image : la SPM.

Des gens ont abouti au modèle linéaire général (*general linear model*, GLM) pour estimer quelques paramètres qui pourraient expliquer la continuité spatiale des données analysées. Plus de détails ont été fournis précédemment. Notez que l'analyse des effets du hasard (*random-effects analysis*) est un sujet différent de ce que nous voyons ici : cette analyse permet de créer des inférences au sujet d'une population à partir de laquelle les sujets ont été tirés et à partir des données récoltées sur ces sujets uniquement (et pas toute la population). Dans le cas de la RFT, nous allons comparer des voxels, pas des sujets.

On en arrive finalement à l'inférence, où la théorie des champs de hasard entre en jeu pour résoudre le problème de comparaison multiple qui survient lorsqu'on doit produire des inférences sur un volume de cerveau.

1.2 Quelques définitions

Pour fixer les idées et essayer de comprendre de quoi on parle, voici quelques définitions générales (généralistes) des mots les plus employés :

- L'**inférence** est une opération logique par laquelle on admet une proposition en vertu de sa liaison avec d'autres propositions déjà tenues pour vraies (synonyme de déduction, d'induction)
- Une **théorie** est une construction intellectuelle méthodique et organisée, de caractère hypothétique (au-moins en certaines de ses parties) et synthétique
- Un **champ** (*field* en anglais) devrait être pris, ici, je crois, comme un domaine (espace ou milieu naturel) où se manifeste un phénomène physique déterminé en tout point
- Finalement, le **hasard** (*random* en anglais) est un cas, un événement fortuit, un concours de circonstances inattendus et inexplicables

1.3 Quels sont les problèmes ?

Dans cette partie, nous nous intéressons à l'inférence au niveau d'un ensemble de voxels. Devant un champ de voxels, cette partie va nous permettre de répondre à la question : “est-ce que l'activation à un voxel donné est significativement différente de zéro?”. Ce voxel est-il activé, oui ou non ? Si oui, est-ce du au hasard ou est-ce un vrai effet biologique ?

A la fin de la modélisation, nous nous retrouvons avec un grand volume de valeurs statistiques. Comment fait-on pour “trier” tout ça ? Comment fait-on pour trouver un effet quelque part ? Et

¹voxel, contraction de “volume pixel” : la plus petite partie distinguable (en forme de boîte ou cube) d'une image tridimensionnelle ... (voir chapitres précédents)

Degrés de libertés	α		
	0.05	0.02	0.01
30	2.042	2.457	2.750
40	2.021	2.423	2.704
60	2.000	2.390	2.660
120	1.980	2.358	2.617
∞	1.960	2.326	2.576

TAB. 1 – Quelques valeurs de la table de distribution nulle pour t

pourquoi là et pas ailleurs ? Ce problème n'est pas rendu facile du fait qu'il y a plusieurs milliers de voxels et donc plusieurs milliers de valeurs statistiques. En un mot : c'est beaucoup !

2 Le problème de la comparaison multiple

Le problème de la comparaison multiple en imagerie fonctionnelle est donc qu'il y a beaucoup de données à traiter.

2.1 L'hypothèse nulle en statistiques classiques

Lorsque nous calculons une valeur statistique, nous souhaitons souvent savoir si cette valeur représente une évidence convaincante de l'effet que nous recherchons. Pour cela, nous utilisons souvent l'**hypothèse nulle**, c.à.d. l'hypothèse qu'il n'y a pas d'effet. Ensuite, nous la comparons à notre valeur.

Pour effectuer cette comparaison, nous avons besoin d'une **distribution nulle** : la distribution de valeurs statistiques que nous attendrions s'il n'y avait aucun effet. En utilisant cette distribution nulle, nous pouvons estimer la (mal)chance que nos valeurs statistiques soient là par hasard.

Par exemple, nous pouvons voir que notre résultat a 5% de chance de provenir de la distribution nulle (en corrolaire, ce résultat a 95% de chance de ne pas en provenir et donc d'être un effet). Nous décidons alors de rejeter l'hypothèse nulle et, donc, d'accepter l'hypothèse alternative qu'il y a un effet. Mais, les 5% sont toujours là : lorsqu'on rejette l'hypothèse nulle, nous acceptons qu'il y ait 5% de (mal)chance que le résultat n'ait, en fait, aucun effet. Nous acceptons qu'il y ait un "risque" de 5% que l'hypothèse nulle soit quand même vraie.

Ces 5% sont le **taux d'erreur de type I** (*type I error rate*) : la chance (ou le risque) que nous prenons d'être dans l'erreur quand nous rejetons l'hypothèse nulle.

Par exemple, si nous faisons un seul test t de student, nous comparons la valeur de t trouvée à la distribution nulle pour la statistique t. Si nous avons une valeur de t de 2.42 et 40 degrés de liberté, la distribution nulle de t (table 1) nous indique que la probabilité d'observer une valeur plus grande ou égale (à 2.42), s'il n'y a pas d'effet est de 0.02. Dans notre cas, nous pouvons rejeter l'hypothèse nulle avec un risque de 2% pour l'erreur de type I.

2.2 Application de l'hypothèse nulle en imagerie fonctionnelle

La situation est plus compliquée en imagerie fonctionnelle parce que nous avons plusieurs voxels et plusieurs valeurs statistiques associées. Si nous ne savons pas, a priori, où notre effet va se produire dans le cerveau, notre hypothèse doit porter sur le volume entier des statistiques dans le cerveau.

Statistique univariée	→	Imagerie fonctionnelle
1 donnée observée	→	plusieurs voxels
1 valeur statistique	→	famille de valeurs statistiques
<i>type I error rate</i>	→	<i>family-wise error rate</i> , FWE
hypothèse nulle	→	hypothèse nulle liée à la famille

TAB. 2 – Table de correspondance des notations statistiques

La preuve contre l’hypothèse nulle serait alors que l’ensemble du volume observé des valeurs ne provienne pas d’une distribution nulle. Donc, nous nous posons maintenant non plus la question de l’effet d’une valeur mais de plusieurs valeurs. On dit qu’on se pose la question de l’effet sur le volume ou la famille de statistiques sur les voxels.

Le risque d’erreur que nous sommes prêts à accepter devient **taux d’erreur lié à la famille** (*Family-wise Error Rate*, FWE). FWE est donc la probabilité que cette famille de valeurs (associées aux voxels) soit apparue par chance. Nous pouvons ainsi résumer les changements dans la table 2.

2.3 Les méthodes de test de l’hypothèse nulle liée à la famille

Nous pouvons tester l’hypothèse nulle liée à la famille (*family-wise null hypothesis*) de 3 manières :

1. avec un seuil de hauteur (*height threshold*),
2. avec une inférence au niveau d’un ensemble (*set-level inference*),
3. avec une inférence au niveau d’un *cluster* (*cluster-level inference*)

On va voir le seuil de hauteur à la section suivante (section 2.4) ; les deux types d’inférences seront vus dans des chapitres et présentations ultérieures.

2.4 Seuil de hauteur et pouvoir de localisation

Pour tester l’hypothèse nulle liée à la famille, une méthode utile est de chercher toute valeur statistique qui est plus grande que nous pourrions attendre si elles venaient toutes d’une distribution nulle.

Mais cette méthode demande qu’on fixe ce que ça veut dire “plus grand que” : plus grand que quelle valeur ? Cette méthode requière donc qu’on fixe un seuil à appliquer à chaque valeur statistique. Ainsi, pour toute valeur statistique qui sera supérieure à ce seuil, il sera improbable qu’elle soit arrivée par chance (hasard) (figure 2).

On cherche donc un seuil de hauteur. On appelle cette recherche la **fixation du seuil de hauteur** (*height thresholding*).

L’avantage principal de cette recherche est qu’elle a un **pouvoir de localisation** (*localising power*). Ainsi, si on trouve qu’une valeur statistique est plus haute que le seuil, nous pouvons conclure qu’il y a un effet au niveau du voxel (de sa localisation) associé à cette valeur statistique.

Cependant, un seuil de hauteur qui peut “contrôler” un taux d’erreur lié à la famille (FWE) doit tenir compte du nombre de test.

En effet, si on reprend l’exemple vu précédemment : une valeur statistique t de la distribution nulle avec 40 degrés de libertés a une probabilité de 2% d’être plus grande que 2.42. Maintenant, imaginons que notre expérience a généré 1000 valeurs de t (toujours avec 40 degrés de libertés). Si nous regardons chacune de ces valeurs, elles auront chacune, individuellement, une probabilité de 2% d’être plus grande que 2.42. Cela signifie que nous nous attendons à avoir 20 valeurs de t (de notre échantillon de 1000 valeurs) plus grandes que 2.42.

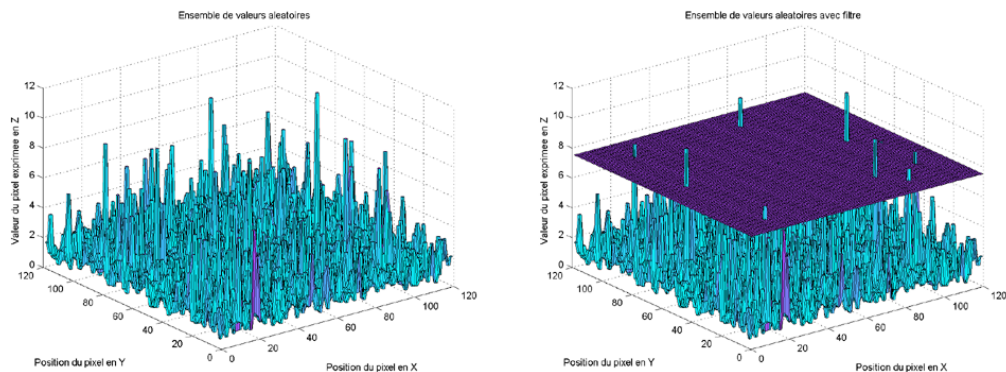


FIG. 2 – A gauche : représentation 3D d’une tranche en utilisant des nombres aléatoires issus d’une distribution normale. La valeur de chaque pixel est traduite en valeur selon l’axe z. A droite : application d’un seuil arbitraire sur cette tranche; dans ce cas, seuls les pixels dont les valeurs dépassent ce seuil seraient reprises.

Mais ce dont on a besoin est une statistique qui ne s’applique pas qu’à un simple voxel, mais simultanément à l’ensemble des voxels concernés (compte-tenu qu’un t-test à n degrés de libertés a été réalisé chez chacun d’entre eux). Lorsqu’on fixe individuellement un seuil de hauteur à chaque pixel, la vue globale ne dit pas dans quelle mesure la famille de valeurs a été observée par chance (càd. l’hypothèse nulle en *family-wise error*) ou pas.

Ainsi, si nous voyons une ou plusieurs valeurs de t au-dessus de 2.42 dans cette famille de tests, suite à une fixation individuelle de seuil de hauteur, ce n’est pas une bonne preuve contre l’hypothèse nulle liée à la famille (qui stipule que toutes / un ensemble de ces valeurs ont été prises dans une distribution nulle).

Nous devons donc trouver un nouveau seuil de telle manière que, dans une famille de 1000 valeurs statistiques t , il y ait une probabilité de 2% qu’il n’y ait qu’une ou plusieurs valeurs de t au-dessus de ce seuil (au niveau de la famille, plus au niveau d’un voxel).

3 La correction de Bonferroni

La correction de Bonferroni ² est la solution “facile”, usuelle au problème de la comparaison multiple, càd. pour déterminer un seuil.

Le but de cette correction est de trouver un seuil pour tester l’hypothèse nulle avec une meilleure méthode que le seuil de hauteur en ce qui concerne des familles de valeurs. La correction de Bonferroni est donc une méthode (simple) de déterminer le seuil au-dessus duquel les valeurs ne sont probablement pas arrivées par chance. Elle est basée sur des règles simples de probabilité.

Imaginez que nous avons pris nos valeurs de t et que nous avons utilisé la distribution nulle pour t pour les convertir en valeurs de probabilité. Nous appliquons alors un seuil de probabilité α à chacune de nos n valeurs de probabilité. Dans notre exemple précédent (section 2.4), α était de 0.02 et n était égal à 1000.

²Carlo Emilio Bonferroni (1892-1960) était un professeur de mathématiques financières en Italie (figure 3)



FIG. 3 – Carlo Emilio Bonferroni

3.1 Expression mathématique de la correction de Bonferroni

Revenons un moment au niveau d'un voxel. Si toutes les valeurs de test sont issues d'une distribution nulle, alors chacune de nos n valeurs de probabilité possède une probabilité α d'être plus grande que le seuil. Comme corollaire, chacune de nos n valeurs de probabilité possède une probabilité $(1 - \alpha)$ d'être inférieure au seuil.

Retournons à la famille de (n) valeurs, la probabilité que *tous* les n tests soient inférieurs à α est donc :

$$(1 - \alpha)^n$$

Nous allons alors définir le **taux d'erreur lié à la famille** (*family-wise error rate*, P^{FWE}) comme la probabilité qu'une ou plusieurs valeurs seront plus grandes que α . En forme mathématique, cela donne :

$$P^{FWE} = 1 - (1 - \alpha)^n \quad (1)$$

Puisque α est petit (0.01, par exemple), on peut étendre cette équation et le négliger en certains endroits ...

$$P^{FWE} = 1 - (1^n - n\alpha + \alpha^n)$$

$$P^{FWE} = 1 - (1 - n\alpha + \alpha^n)$$

$$P^{FWE} = 1 - 1 + n\alpha - \alpha^n$$

$$P^{FWE} = n\alpha - \alpha^n$$

On peut donc affirmer que, approximativement (α^n est pratiquement nul puisque α est très petit),

$$P^{FWE} \leq n \cdot \alpha \quad (2)$$

En utilisant cette inéquation, nous pouvons trouver α , le seuil de probabilité pour un seul voxel, qui va nous donner le taux d'erreur lié à la famille (P^{FWE}) tel que nous aurons une probabilité P^{FWE} de voir n'importe quel voxel au-dessus du seuil dans toutes les n valeurs.

On peut résoudre simplement cette inéquation pour α :

$$\alpha = \frac{P^{FWE}}{n} \quad (3)$$

	1 valeur	famille de valeurs
Nombre de valeurs statistiques	n (100 000)	
Nombre de degrés de libertés	(40)	
Taux d'erreur p	α	P^{FWE}
p corrigé pour la famille ?	non	oui

TAB. 3 – Table de correspondance des notations statistiques et abbréviations

3.2 Exemple

Nous avons un volume de cerveau de 100 000 ($= n$) valeurs statistiques de t , toutes avec un degré de liberté de 40. Nous voulons un taux d'erreur lié à la famille (*FWE rate*) de 0.05 ($= P^{FWE}$). Prenons la correction de Bonferroni et calculons le seuil de probabilité requis pour un seul voxel. Ce seuil (α) sera de $\alpha = \frac{0.05}{100000} = 0.000005$. La statistique t correspondante sera 5.77.

Ainsi, si n'importe quelle statistique t de voxel est supérieure à 5.77, nous pouvons alors conclure qu'une statistique de voxel de cette amplitude a seulement 5% de chance de se trouver dans un volume de 100 000 valeurs statistiques t tirés d'une distribution nulle (autrement dit : il y a beaucoup de chance que cette valeur ne soit pas due au hasard).

La procédure de Bonferroni donne une valeur corrigée de p (*corrected p value*) : c'est P^{FWE} et sa valeur est de 0.05. Dans le cas ci-dessus, la valeur de p non corrigée pour un voxel (α) avec une valeur statistique t de 5.77 était de 0.000005. La table 3 résume ces différences.

3.3 La correction de Bonferroni n'est souvent pas applicable !

La correction de Bonferroni est utilisée pour calculer le taux d'erreur lié à la famille (*FWE rate*) de quelques analyses d'imagerie fonctionnelle.

Malheureusement, dans la plupart des autres cas ³, la correction de Bonferroni est trop conservative. On a jusqu'à présent considéré qu'il y avait autant de valeurs indépendantes (dans le volume de valeurs statistiques) qu'il y avait de voxels. Or, la plupart des données d'imagerie fonctionnelle ont un certain degré de corrélation spatiale : il y a une certaine corrélation entre les valeurs statistiques voisines. Il y a donc moins de valeurs réellement indépendantes qu'il n'y a de voxels.

On va revenir sur les raisons pour lesquelles la correction de Bonferroni n'est pas applicable dans la plupart des études d'imagerie fonctionnelle (section 5). Mais, avant cela, attardons-nous un peu sur la corrélation spatiale.

4 La corrélation spatiale

En général, les données de n'importe quel voxel d'une image fonctionnelle tendra à être similaire aux données de ses voxels voisins, même après que les effets modélisés aient été enlevés. C'est pour cela qu'on dit qu'un certain nombre de degrés de corrélation est toujours présent dans les données d'imagerie fonctionnelle.

Donc, les erreurs provenant du modèle statistique vont tendre à être aussi corrélées aux voxels voisins. Les raisons à cela sont multiples : il y a

- des facteurs inhérents à la collecte des images
- le signal physiologique qui n'a pas été modélisé (voir chapitres précédents)
- le pré-traitement spatial appliqué aux données avant analyse statistique

Nous nous étendrons sur deux de ces raisons.

³Quels cas en particuliers ? Demandez à des spécialistes !

4.1 Corrélation due à la collecte des données

Dans les données de PET (*Positron Emission Tomography*), plus que dans celles de fMRI, les voxels voisins sont liés entre eux à cause de la manière dont le scanner collecte et reconstruit l'image. Ainsi, les données qui proviennent en fait d'un seul voxel vont aussi changer, d'une certaine manière (jusqu'à un certain degré), le signal dans les voxels voisins dans l'image résultante. L'étendue de cette influence est une mesure de la performance d'un scanner PET. On en parlera comme la *point spread function*.

4.2 Corrélation due au pré-traitement

Le pré-traitement spatial des données fonctionnelles introduit également une corrélation spatiale. Les pré-traitements sont les ré-alignements spatial des images, le lissage, etc.

4.2.1 Ré-alignement, normalisation spatiale, ré-échantillonnage

Par exemple, cela se passe quand on ré-aligne des images d'un seul sujet pour corriger le mouvement du corps durant la session de scan (voir le chapitre 2 du livre [1]) ou quand on normalise spatialement le cerveau d'un sujet sur un modèle (*template*) pour pouvoir comparer les données entre les sujets (voir le chapitre 3 du livre [1]).

En effet, ces transformations requièrent la création de nouvelles images ré-échantillonnées dont les centres de voxel sont très probablement différent que ceux dans les images originales. Le ré-échantillonnage exige que nous estimons le signal dans ces nouveaux voxels à partir des valeurs dans l'image originale et les méthodes courantes de ré-échantillonnage demandent un certain degré de "moyennage" des voxels voisins pour dériver la nouvelle valeur d'un voxel (voir le chapitre 2 du livre [1]).

4.2.2 Lissage

Il est aussi très commun de lisser (*to smooth*) les images fonctionnelles avant l'analyse statistique. Une proportion du bruit dans les images fonctionnelles est indépendante de voxel en voxel alors que le signal d'intérêt s'étend généralement sur plusieurs voxels. Cela est dû à deux facteurs (voir le chapitre 11 du livre [1]) :

1. la nature potentiellement distribuée des sources neuronales
2. la nature étendue dans l'espace de la réponse hémodynamique

Selon le théorème du filtre assorti (*matched filter theorem*), le lissage va donc améliorer le rapport signal sur bruit. Pour les analyses de nombreux sujets, le lissage peut également être utile pour brouiller (*to blur*) les différences résiduelles de localisation entre les aires correspondantes d'activation fonctionnelle. Le lissage implique un "moyennage" sur les pixels, ce qui, par définition, va augmenter la corrélation spatiale.

Nous allons voir 2 exemples de lissage à la section suivante.

5 Bonferroni n'est pas applicable en imagerie fonctionnelle

Donc, nous avons dit que la correction de Bonferroni n'est généralement pas applicable en imagerie fonctionnelle, à cause de la corrélation spatiale des données. En effet, la corrélation spatiale signifie qu'il y a moins d'observations indépendantes dans les données qu'il n'y a de voxels.

5.1 Un peu de maths

La probabilité liée à la famille de l'équation 1 repose sur le fait que les valeurs de probabilité individuelles sont indépendantes. De cette manière, nous avons pu utiliser la multiplication pour calculer la probabilité que tous les tests seront sous le seuil avec $(1 - \alpha)^n$. Ainsi, le n dans l'équation doit être pris comme le nombre d'observations *indépendantes*.

Si, maintenant, nous avons bien n voxels dans nos données mais seulement n_i observations indépendantes (avec $n \leq n_i$), l'équation 1 devient :

$$P^{FWE} = 1 - (1 - \alpha)^{n_i}$$

Et le α correspondant à l'équation 3 est donné par :

$$\alpha = \frac{P^{FWE}}{n_i}$$

Comme d'habitude, c'est plus parlant avec un exemple.

5.2 Exemple

Prenons une seule tranche, de 100 pixels sur 100, avec un test statistique t pour chaque voxel. Pour faire simple, donnons à ces valeurs statistiques t un grand degré de liberté ; de cette manière, nous pourrions considérer que les valeurs statistiques t sont issues d'une distribution normale, c.à.d. qu'ils sont des scores Z [C'EST QUOI???] (on verra plus tard qu'on peut appliquer d'autres distributions, cfr. section 6.4). Nous pouvons simuler cette tranche à partir d'une distribution nulle en remplissant les valeurs des voxels avec des nombres aléatoires indépendants de la distribution normale, ce qui va résulter en une image semblable à la figure 4.

Si cette image provenait de l'analyse de données réelles, nous pourrions vouloir tester si n'importe lequel de ces nombres dans l'image est plus positif qu'il le serait par simple chance. Les valeurs sont indépendantes ; donc, la correction de Bonferroni donnera un seuil précis.

Il y a 10 000 scores Z ($100 \cdot 100 = 10000$). Pour un P^{FWE} de 0.05, le seuil corrigé de Bonferroni pour α est de $\frac{0.05}{10000} = 0.000005$. Cela correspond à un score Z de 4.42.

Etant donné l'hypothèse nulle (qui est vraie dans ce cas), nous nous attendons à ce que seulement 5 images sur 100 images semblables possèdent un ou plusieurs scores Z plus positif que 4.42 par chance.

Mais la situation change si nous ajoutons une corrélation spatiale ! Maintenant, nous allons effectuer une simple opération sur l'image (l'image résultante est à la figure 5) :

1. coupons l'image en carrés de 10 pixels sur 10
2. pour chaque carré, calculons la moyenne des 100 valeurs contenues
3. remplaçons les 100 nombres aléatoires dans le carré par la valeur moyenne ⁴

Nous avons toujours 10 000 nombres dans notre image. Mais il n'y a que $10 \cdot 10 = 100$ nombres indépendants. La correction appropriée de Bonferroni est maintenant de $\frac{0.05}{100} = 0.0005$, ce qui correspond à un score Z de 3.29. Ainsi, maintenant, nous nous attendons à ce que seulement 5 images sur 100 images semblables possèdent un carré de valeurs plus grandes que 3.29 par chance.

Si nous avions supposés que toutes les valeurs étaient indépendantes, nous aurions utilisé la correction pour 10 000 valeurs, avec $\alpha = 0.000005$. Mais parce que nous n'avons maintenant que

⁴Notons que "moyenner" des nombres aléatoires va les faire tendre vers zéro. Pour que l'image revienne à une variance de 1, nous devons multiplier ces nombres par 10. En fait, ce nombre 10 n'est pas du au hasard : $10 = \sqrt{100}$. Si on généralise, le nombre à multiplier est \sqrt{n} , où n est le nombre de valeurs que nous avons "moyennées" dans chacun des carrés.

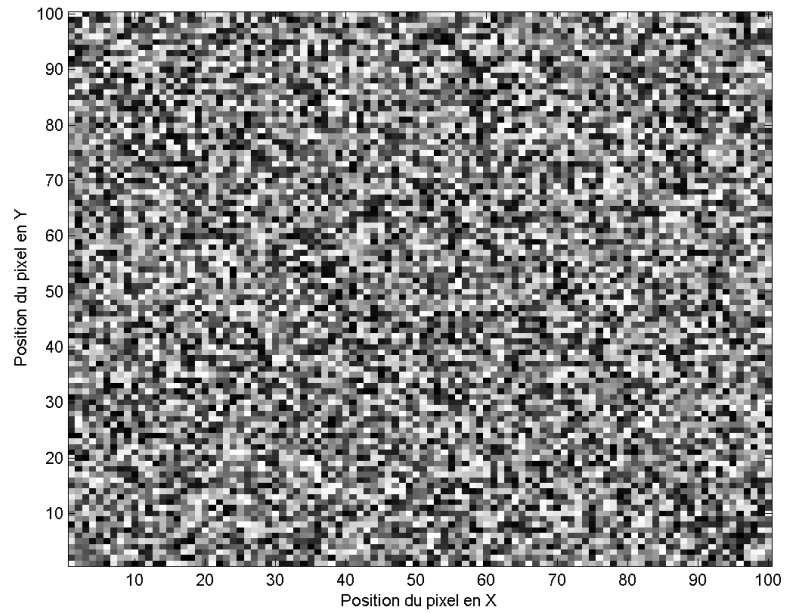


FIG. 4 – Image simulée d’une tranche en utilisant des nombres aléatoires issus d’une distribution normale. Les pixels blancs sont les plus positifs

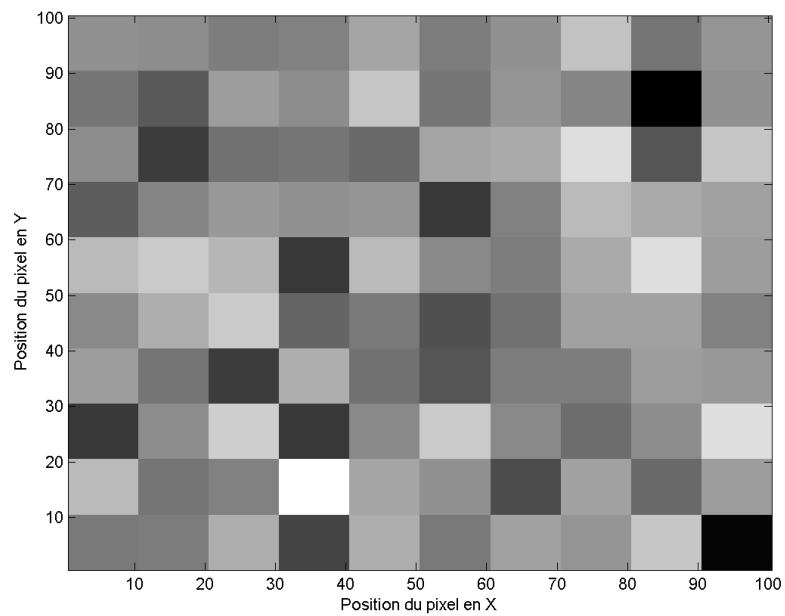


FIG. 5 – Image des nombres aléatoires de la figure 4 après avoir remplacé les valeurs dans des carrés de 10 pixels sur 10 par la valeur de la moyenne de chacun de ces carrés

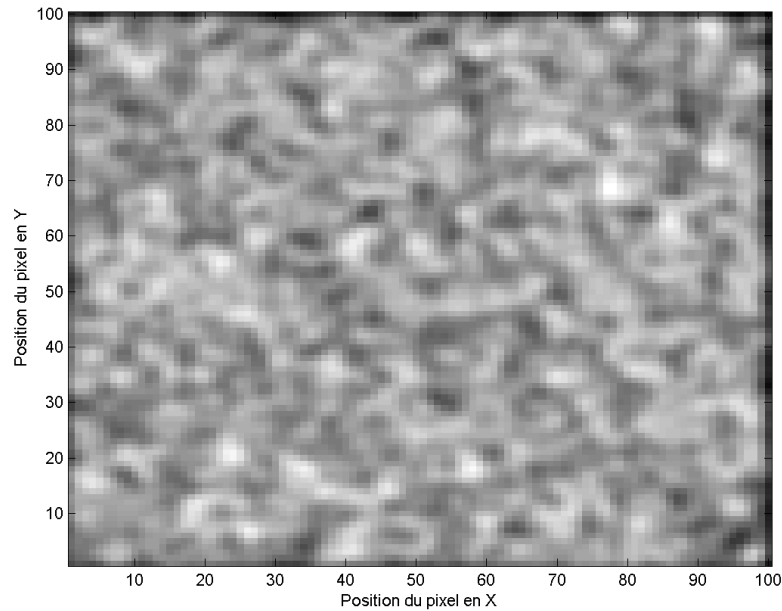


FIG. 6 – Image des nombres aléatoires de la figure 4 après lissage par Gauss

100 observations indépendantes, remplaçons $n = 100$ et $\alpha = 0.000005$ dans l'équation 2, nous verrons que nous obtenons un taux de FWE (P^{FWE}) de 0.0005. C'est 100 fois plus petit (càd. plus conservateur) que le taux que nous voulions.

5.3 Le lissage et les observations indépendantes

Dans la section précédente, on remplaçait un carré de valeurs avec leur moyenne pour montrer l'effet de la réduction du nombre de variables indépendantes. Cette procédure est une forme très simple de lissage. Augmentons encore la difficulté en employant une autre forme de lissage, plus complexe ...

Lorsque nous lisons une image avec un noyau de lissage comme un noyau gaussien (*Gaussian kernel*), par exemple, chaque valeur de l'image est remplacé par une moyenne pondérée (*weighted average*) d'elle-même et de ses voisines. Par exemple, la figure 6 montre l'image de la figure 4 après lissage par un noyau gaussien dont le *Full Width at Half Maximum* (FWHM) est de 10 pixels⁵. Un FWHM de 10 pixels signifie que, à 5 pixels du centre, la valeur du noyau est égale à la moitié de sa valeur totale.

De nouveau, le lissage a pour effet de brouiller l'image et il réduit le nombre d'observations indépendantes. L'image lissée contient donc une corrélation spatiale, ce qui est typiquement ce qu'on retrouve en sortie d'une analyse des données d'imagerie fonctionnelle.

Maintenant, nous avons un problème : comment trouver le seuil ? En effet, dans le lissage

⁵Comme dans la procédure où nous prenons la moyenne de 100 observations dans chacun des carrés, les valeurs lissées n'auront plus une variance de 1, de nouveau parce que le "moyennage" impliqué dans ce lissage fait tendre les valeurs vers zéro. Comme dans l'exemple des carrés, nous devrons multiplier les valeurs de l'image lissée par un facteur d'échelle pour rétablir cette variance à 1. Cependant, la méthode pour trouver ce facteur est très technique et n'est pas très utile dans notre discussion

donnant la figure 5, nous connaissions le nombre de valeurs indépendantes; ce qui nous a permis de calculer le seuil. Mais, dans les valeurs lissées de la figure 6, il n'est pas facile de calculer le nombre de variables indépendantes. Donc, nous ne pouvons pas utiliser la correction de Bonferroni. Que faire? Tel Zorro, la *Random Field Theory* vient au secours du "FRMIen" ...

6 La Random Field Theory

Utilisée avec sagesse (càd. après avoir bien compris les suppositions préalables), la **théorie de champ de hasard** (*Random Field Theory*, RFT) est la bonne solution au problème de la comparaison multiple. Elle va permettre de trouver le seuil dans un ensemble de données où il n'est pas facile (voire impossible) de calculer le nombre de variables indépendantes.

La RFT est un corps récent des mathématiques définissant les résultats théoriques pour des cartes statistiques lissées. Cette théorie est assez souple pour s'attaquer à de nombreux problèmes de seuil que nous rencontrons en imagerie fonctionnelle. Entre autres applications, elle peut être utilisée pour résoudre notre problème pour trouver le seuil de hauteur dans une carte statistique lissée qui donne le taux d'erreur lié à la famille requis.

Pour résoudre ce problème, la RFT utilise les résultats donnés par la **caractéristique d'Euler** (EC) attendue (*expected Euler characteristic*) pour une carte statistique lissée à laquelle on a appliqué un seuil.

On va discuter plus en détails de la caractéristique d'Euler plus tard (section 6.2). Pour le moment, il suffit de noter que la caractéristique d'Euler attendue mène directement au nombre attendu de clusters (groupes) au-dessus d'un certain seuil et que, à son tour, ce nombre donne le seuil de hauteur dont nous avons besoin :

$$\text{EC attendue} \rightarrow \text{nombre de clusters au-dessus du seuil} \rightarrow \text{seuil de hauteur}$$

L'application de la RFT procède par étapes :

1. D'abord, nous estimons le caractère lisse (*smoothness*; la corrélation spatiale) de notre carte statistique (voir point 6.1)
2. Ensuite, nous utilisons ces valeurs de lissage dans l'équation appropriée de la RFT; ce qui nous donne l'EC attendue à différents seuils (voir point 6.2)
3. Finalement, cela nous permet de calculer le seuil auquel nous nous attendons à ce que 5% des cartes statistiques équivalentes issues de l'hypothèse nulle contiennent au-moins une aire au-dessus du seuil (voir point 6.2.4).

6.1 Le caractère lisse et les resels

D'habitude, nous ne connaissons pas le caractère lisse (*smoothness*) de notre carte de valeurs statistiques. Et c'est le cas même si la carte résulte de données lissées car nous ne connaissons pas, d'habitude, l'étendue de la corrélation spatiale dans les données sous-jacentes avant lissage.

Même si nous ne connaissons pas ce caractère lisse, il peut être calculé en utilisant la corrélation spatiale observée dans les images. Nous verrons ce cas à la section 6.4 : c'est ce cas qui est utilisé pratiquement en imagerie fonctionnelle.

Dans l'exemple de la figure 6 cependant, nous connaissons la caractère lisse car les données étaient indépendantes avant lissage. Dans ce cas, le caractère lisse résulte entièrement du lissage que nous avons appliqué.

Ce caractère lisse peut être exprimé comme la largeur du noyau de lissage, qui était de 10 pixels FWHM dans les directions x et y. Nous pouvons utiliser le FWHM pour calculer le nombre de resels dans l'image.

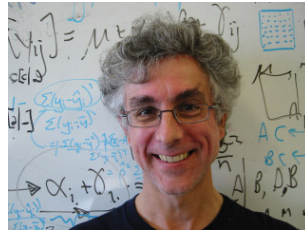


FIG. 7 – Keith John Worsley

Resel est un terme introduit par Keith Worsley ⁶. Un resel est une mesure du nombre d’“éléments de résolution” (*resolution elements*) dans une carte statistique. On pourrait voir cela comme le nombre d’observations indépendantes mais ce n’est pas la même chose (voir section 6.3).

Un resel est défini simplement comme un bloc de valeurs (dans notre cas, des pixels) qui est de même taille que le FWHM.

Pour l’image à la figure 6, les FWHMs étaient de 10 pixels sur 10. Donc, un resel est un bloc de 100 pixels. Puisqu’il y a 10 000 pixels dans notre image, il y a 100 resels.

Notons que le nombre de resels ne dépend que du caractère lisse (FWHM) et du nombre de pixels.

6.2 La caractéristique d’Euler

La **caractéristique d’Euler** (EC) est une propriété d’une image après qu’on y ait appliqué un seuil. Dans notre cas, l’EC peut être vue comme le nombre de gouttes (*blobs*) dans une image après application d’un seuil.

6.2.1 Exemple

Dans notre exemple, nous pouvons appliquer un seuil à $Z = 2.5$ à la figure 1 du livre : tous les pixels dont les scores Z sont inférieurs à 2.5 sont défini à zéro et le reste à 1. Le résultat est à la figure 8.

Il y a trois gouttes blanches dans la figure 8, correspondant aux trois surfaces où les valeurs de Z sont supérieures à 2.5. L’EC de cette image est donc 3.

Si nous augmentons le seuil de score Z à 2.75, nous voyons que les deux gouttes centrales disparaissent (parce que leurs scores Z étaient supérieurs à 2.5 mais inférieurs à 2.75 ; figure 10). La surface dans le coin supérieur droit de l’image reste. L’EC de la figure 10 est donc de 1 (un autre exemple de l’application de ces différents seuils est montré à la figure 9).

6.2.2 Signification de la caractéristique d’Euler attendue

A des seuils élevés, l’EC est soit 1 soit 0. Par conséquent, l’EC moyenne ou attendue, notée $E[EC]$, correspond approximativement à la probabilité de trouver une goutte à seuil plus élevé dans notre image statistique. Ainsi, la probabilité d’une erreur liée à la famille est approximativement équivalente à l’EC attendue :

$$P^{FWE} \approx E[EC]$$

⁶Keith John Worsley est un professeur au département de mathématiques et statistiques de l’Université de McGill, au Canada (figure 7). Il est notamment impliqué dans le projet FMRISTAT (des outils sous Matlab pour l’analyse statistique générale de données issues de la fMRI)

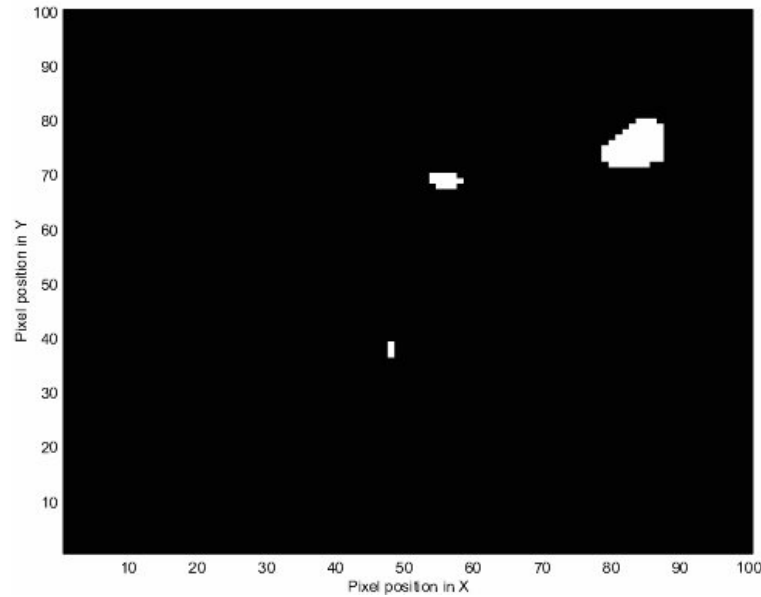


FIG. 8 – Image lissée des nombres aléatoires de la figure 6 après application d’un seuil de $Z = 2.5$. Les valeurs inférieures au seuil deviennent des zéros (affichées en noir). Les valeurs restantes deviennent des 1 (affichées en blanc) (image reprise du livre)

6.2.3 Calcul de la caractéristique d’Euler attendue

Il s’avère que, si nous connaissons le nombre de resels dans notre image, il est possible de calculer la $E[EC]$ pour n’importe quel seuil. Par exemple, pour une image de 2 dimensions, $E[EC]$ est donné dans le papier de Worsley [2] (pas lu, bien sûr!) ... Si R est le nombre de resels et que Z_t est le seuil de score Z , alors :

$$E[EC] = R \cdot (4 \log_e 2) \cdot (2\pi)^{-\frac{2}{3}} \cdot Z_t \cdot e^{-\frac{1}{2}Z_t^2} \quad (4)$$

La figure 11 ⁷ montre $E[EC]$ pour une image de 100 resels, pour des scores Z entre 0 et 5.

Lorsque le seuil tombe de 1 à 0, $E[EC]$ tombe vers zéro aussi; c’est parce que la définition précise de la EC est plus complexe que simplement le nombre de gouttes. Cela fait une différence aux seuils bas mais n’est pas trop intéressant dans notre cas parce que, comme expliqué ci-dessus, nous ne sommes intéressés qu’aux propriétés de $E[EC]$ à des seuils élevés, càd. lorsqu’elle vaut approximativement P^{FWE} .

Notons aussi que le graphique à la figure 11 donne une bonne approximation de la EC de notre image ⁸. A un seuil de $Z = 2.5$, il prédit une EC de 1.9 alors que nous avons observé une valeur de 3. A un seuil de $Z = 2.75$, il prédit une EC de 1.1 pour une EC observée de 1.

6.2.4 Application de la caractéristique d’Euler calculée à notre exemple

Nous pouvons maintenant appliquer la RFT à notre image lissée (notre figure 6) qui possède 100 resels.

⁷Je n’ai pas trouvé pourquoi mon application de la formule de $E[EC]$ me donnait un maximum aux environs de 49 alors que la figure 6 du livre donne un maximum aux environs de 11.

⁸Il faut tenir compte de la note précédente. La phrase suivante est la copie du livre, pas l’observation de la figure.

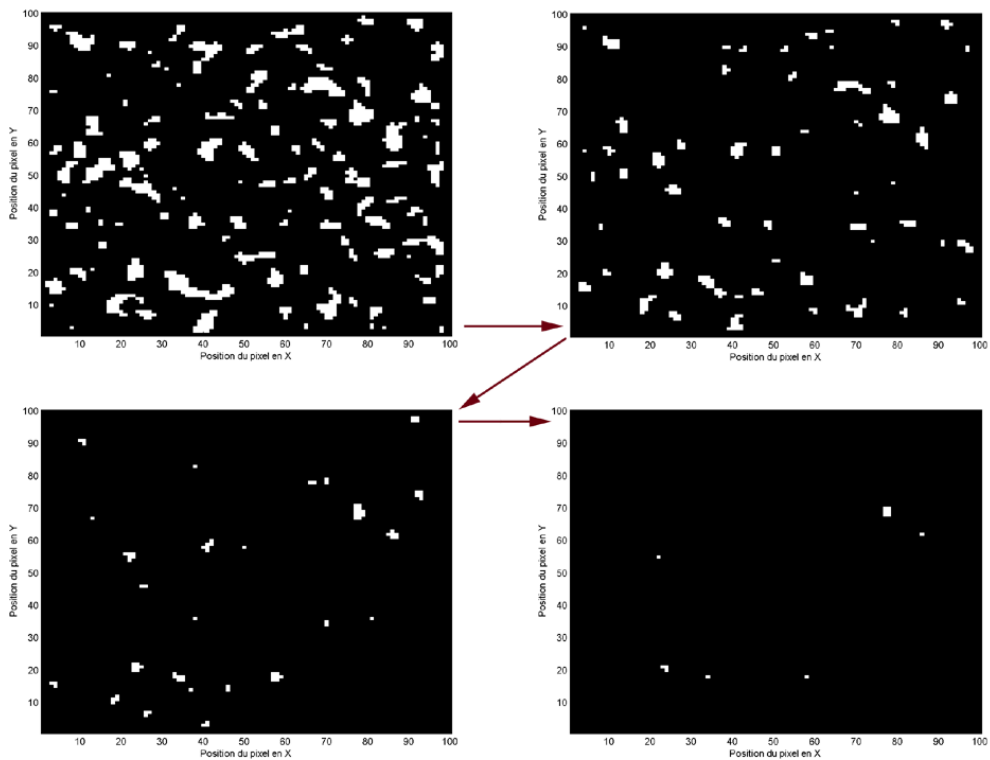


FIG. 9 – Image lissée des nombres aléatoires de la figure 6 après application d'un seuil croissant (sens des flèches). Les valeurs inférieures au seuil deviennent des zéros (affichées en noir). Les valeurs restantes deviennent des 1 (affichées en blanc)

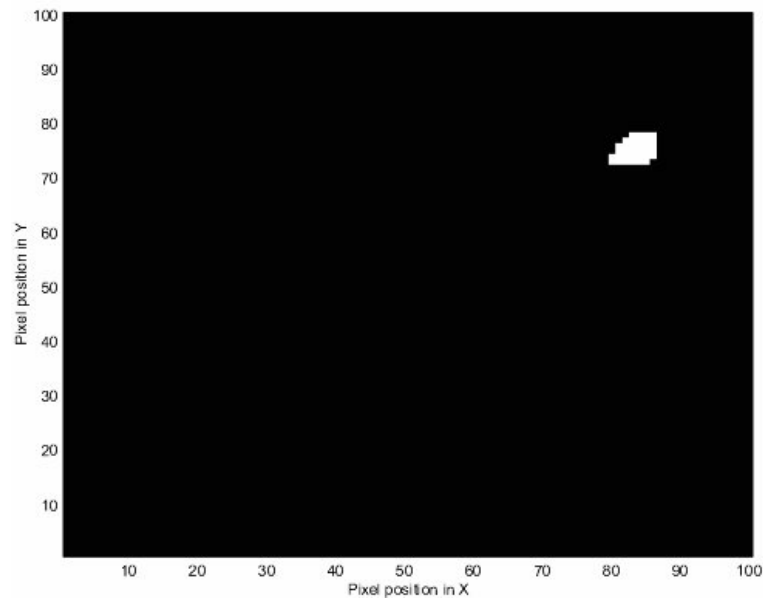


FIG. 10 – Image lissée des nombres aléatoires de la figure 6 après application d’un seuil à $Z = 2.75$. Les valeurs inférieures à 2.5 deviennent des zéros (affichées en noir). Les valeurs restantes deviennent des 1 (affichées en blanc) (image reprise du livre)

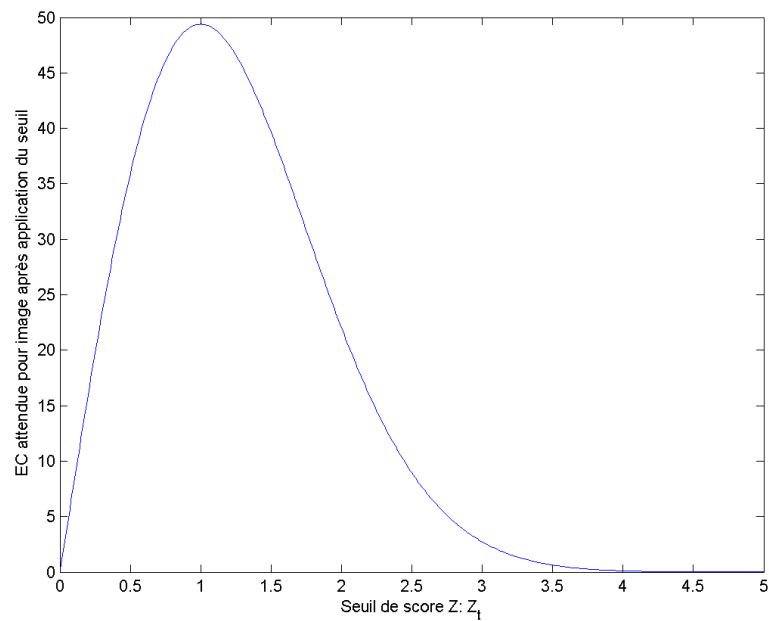


FIG. 11 – Valeurs de la caractéristique d’Euler attendue pour une image de 100 reseles et calculées avec l’équation 4

α	number de resels dans l'image	Bonferroni		RFT
		seuil	score Z	seuil Z
0.05	100	$\frac{0.05}{100}$	3.3	
				3.8

TAB. 4 – Comparaison des réponses données par la RFT et Bonferroni

Pour 100 resels, l'équation 4 donne une $E[EC]$ de 0.0049 pour un seuil Z de 3.8 (voir le graphe de la figure 11). Si nous avons une image bi-dimensionnelle avec 100 resels, alors la probabilité d'avoir une ou plusieurs gouttes plus grandes que 3.8 est 0.049. Nous pouvons utiliser cette valeur pour déterminer le seuil.

Posons x comme le seuil de score Z qui donne une $E[EC]$ de 0.05. Si nous appliquons un filtre de x à notre image, nous pouvons conclure que toutes gouttes qui pourraient subsister à une probabilité inférieure ou égale à 0.05 qu'elles soient là par chance, hasard. Si on reprend l'équation 4, ce seuil x ne dépend que du nombre de resels dans notre image.

6.3 Les seuils trouvés par la RFT et la correction de Bonferroni

La correction (*random field correction*) trouvée par la RFT n'est pas la même que la correction de Bonferroni pour le même nombre de resels. Nous avons en effet dit, plus haut (section 6.1), que le nombre de resels dans une image n'est pas exactement le même que le nombre d'observations indépendantes.

S'il était le même, à la place d'utiliser la RFT, nous pourrions utiliser une correction de Bonferroni basée sur le nombre de resels (ce serait plus simple à calculer). Cependant, ces deux corrections donnent des réponses différentes (voir table 4) :

- Pour $\alpha = 0.05$, le seuil de Z selon la RFT, pour notre image de 100 resels, est de 3.8
- Le seuil de Bonferroni pour 100 tests indépendants est de $\frac{0.05}{100}$, ce qui équivaut à un score Z de 3.3

Ainsi, nous voyons que, bien que l'exposé mathématique de la RFT nous donne une correction qui, dans son principe, est similaire à celle de Bonferroni, ce n'est pas la même chose. Si les suppositions de la RFT sont rencontrées (voir section 7), alors le seuil donné par la RFT est plus précis que celui donné par Bonferroni.

6.4 La RFT et l'imagerie fonctionnelle 3D

Maintenant, on quitte les modèles théoriques bien beaux et on entre dans la réalité. Première différence : on a des volumes 3D et plus des cartes 2D !

Les analyses d'imagerie fonctionnelle utilisent généralement des images de statistiques tridimensionnelles. Jusqu'à présent, nous avons parlé de l'application de la RFT à une image en 2 dimensions ; les mêmes principes s'appliquent en 3 dimensions :

- La EC est le nombre de gouttes 3D de scores Z au-dessus d'un certain seuil
- Un resel est un cube de voxels de taille
 - FWHM en x,
 - FWHM en y,
 - FWHM en z,
- L'équation pour $E[EC]$ est différente dans le cas de la 3D mais elle dépend toujours uniquement des resels dans l'image.

Notons que, par simplicité, nous avons seulement considéré un RF de scores Z (càd. nombres issus d'une distribution normale). Mais il existe des résultats équivalents pour les RF de scores t ,

F et χ^2 . C'est notamment utilisé dans SPM99.

Comme nous l'avons indiqué dans la section 6.1, nous ne connaissons généralement pas le caractère lisse d'un volume statistique issu de l'imagerie fonctionnelle. Simplement parce que nous ne connaissons pas l'étendue de la corrélation spatiale avant lissage. Nous ne pouvons pas supposer que le caractère lisse est le même que n'importe quel lissage que nous avons appliqué volontairement. Nous devons donc calculer ce caractère lisse à partir des images elles-mêmes (ce sont les seules données que nous avons). En pratique, le caractère lisse (*smoothness*) est calculé en utilisant les valeurs résiduelles de l'analyse statistique (c'est expliqué dans des références citées dans le livre).

6.5 Les vraies dépendances de l'EC attendue

Deuxième "coup dur" aux modèles théoriques : la forme et le volume prennent maintenant toute leur importance!

On a dit, plus haut (section précédente), que les résultats de la caractéristique d'Euler attendue dépendait seulement du nombre de resels contenus dans le volume de voxels analysé. Ce n'est pas exactement précis bien que ce soit une approximation très proche de la réalité lorsque le volume de voxels est large comparé à la taille d'un resel. En fait, $E[EC]$ dépend aussi de la forme et la taille du volume.

La forme du volume devient importante lorsque nous avons une petite région ou une région curieusement formée. C'est inhabituel lorsque nous étions le cerveau entier. Mais il existe des situations où nous souhaitons restreindre notre recherche à un sous-ensemble plus petit du volume (par exemple, lorsque nous avons une hypothèse spécifique sur le lieu où le signal devrait apparaître).

La raison de l'influence de la forme du volume sur la correction est mieux visible avec un exemple. Retournons à notre image 2D de nombres au hasard et ayant subi un lissage (figure 6). Nous pourrions imaginer que nous avons des raisons de penser qu'un changement de signal va se produire seulement au centre de l'image. Notre régions de recherche ne sera plus l'image toute entière mais pourrait être une boîte au centre de l'image, de taille 30 pixels sur 30 (figure 12).

La boîte contient 9 resels. La figure montre une grille de marqueurs en forme de X, espacés de la largeur d'un resel partout sur l'image, càd. 10 pixels. La boîte contient un maximum de 16 de ces marqueurs. Maintenant, imaginons que nous ayons une région de forme moins usuelle. Pour une raison quelconque, nous pourrions nous attendre à ce que notre signal d'intérêt se passera dans un cadre de 2.5 pixels de large autour de l'extérieur de l'image. Ce cadre contient le même nombre de voxels que la boîte et possède donc le même volume en terme de voxels. Cependant, le cadre contient beaucoup plus de marqueurs (32), c'est pourquoi le cadre est échantillonné par rapport à des données de plus de resels que la boîte. La correction pour les comparaisons multiples pour la cadre doit donc être plus stricte (stricte) que celle pour la boîte.

En fait (amusons-nous), $E[EC]$ dépend réellement :

- du volume,
- de l'aire de surface et
- du diamètre de la région de recherche.

Ces paramètres peuvent être calculés pour des formes continues pour lesquelles les formules de volume, d'aire de surface et de diamètre existent (par exemple, les sphères ou les boîtes). Sinon, les paramètres peuvent être estimés à partir de toute forme qui aurait été définie dans une image.

Restreindre la région de recherche à un petit volume dans la carte statistique peut mener à des seuils grandement réduits pour des taux de FWE donnés. Pour la figure 13, nous avons supposé une analyse statistique résultant en une carte statistique en t de caractère de lissage de 8mm dans les directions X, Y et Z. La statistique t a 200 degrés de libertés et nous avons une région de recherche sphérique. Le graphique montre la valeur de la statistique t qui donne une valeur corrigée de p de 0.05 pour des sphères de diamètre croissant.

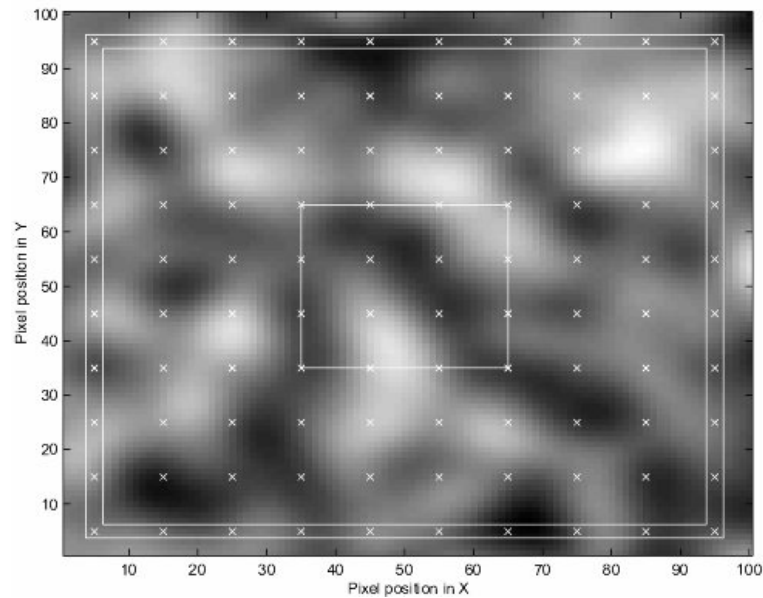


FIG. 12 – Image lissée de nombres aléatoires de la figure 6 avec deux exemples de régions de recherche : une boîte (au centre) et un cadre (bordure externe de l'image). Les marqueurs en forme de X sont espacés de la largeur d'un resel partout sur l'image (image reprise du livre)

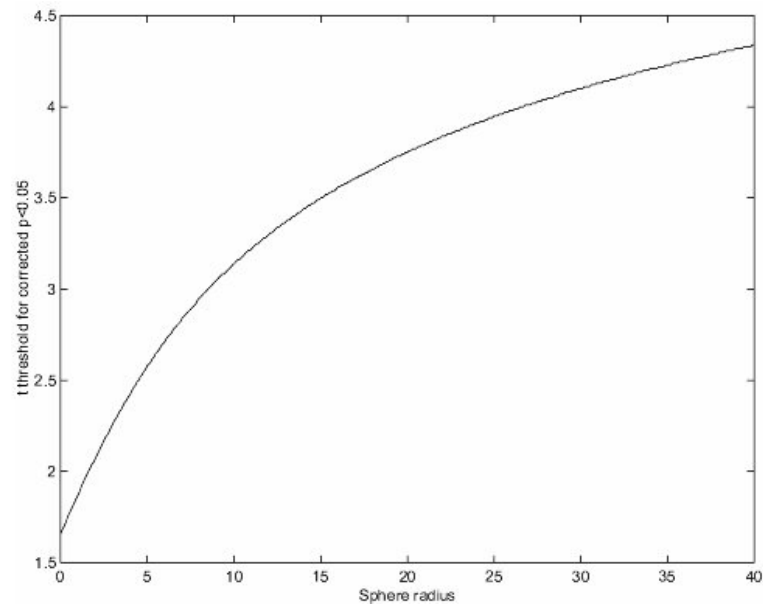


FIG. 13 – Seuil de t donnant un taux de FWE de 0.05 pour des sphères de diamètre croissant. Le caractère lisse est de 8mm dans les directions X, Y et Z. L'exemple d'analyse a 200 degrés de liberté (image reprise du livre)

Analysons cette figure ... Pour une sphère de rayon zéro, le seuil est simplement celui pour une seule statistique t (p non corrigé = p corrigé = 0.05 pour $t = 1.65$ avec 200 degrés de libertés). Le seuil t corrigé augmente brusquement au fur et à mesure que le rayon augmente jusque $\approx 10\text{mm}$ et augmente moins rapidement par après.

6.6 Les hypothèses régionales

Troisième (et dernier) “coup dur” aux modèles théoriques : on ne travaille pratiquement jamais sur tout le volume du cerveau !

En effet, lorsqu’on réalise des inférences à propos d’un effet régional (par exemple, des activations régionales) dans SPM, on a souvent une idée de la localisation approximative où cette activation devrait être. Dans ce cas, une correction pour le volume entier de recherche est inapproprié.

- Si la région hypothétique contient un seul voxel, alors l’inférence pourrait être réalisée en utilisant une valeur de p non corrigée (puisque’il n’y a pas de volume supplémentaire de recherche pour lequel il faut corriger).
- En pratique, cependant, la région hypothétique contiendra d’habitude plusieurs voxels et pourra être caractérisée par des sphères ou des boîtes centrées sur la région d’intérêt (par exemple). Nous devons donc utiliser une valeur p qui a été corrigée de manière appropriée. Comme décrit dans la section précédente, cette valeur corrigée de p sera dépendante de la taille et la forme de la région hypothétique ; elle dépendra aussi du caractère lissé de l’image statistique.

Quelques groupes de recherche ont utilisé des seuils de valeurs de p non corrigées (par exemple $p < 0.001$) de manière à contrôler FWE lorsqu’il y a une hypothèse régionale sur le lieu où l’activation se passerait. C’est mal et il y a un exemple dans le livre.

7 Suppositions préalables

Il y a deux suppositions préalables, sous-jacentes à la RFT :

1. les champs d’erreur (*error fields*) sont une approximation raisonnable de treillis [?] d’un champ de hasard sous-jacent avec une distribution Gaussienne multivariée
2. ces champs sont continus et possède une fonction d’autocorrélation doublement différentiable (*twice-differentiable autocorrelation function*)

Notez qu’une idée fausse généralement répandue est que la fonction d’autocorrélation doit être gaussienne. Ce n’est pas le cas !

Si les données ont été suffisamment lissées et que le modèle linéaire général a été correctement spécifié (de manière à ce que les erreurs soient effectivement gaussiennes), alors les suppositions de la RFT seront rencontrées.

7.1 Cas où les suppositions préalables de la RFT ne sont pas respectées

Un scénario dans lequel les suppositions peuvent ne pas être rencontrées se trouve dans l’analyse des effets de hasard (*random effects analysis*) avec un petit nombre de sujets. Cela résulte du fait que les champs d’erreur résultant ne seront pas très lisses et, ainsi, peuvent être en contradiction avec la supposition d’être un “approximation raisonnable de treillis”.

Une solution est de réduire la taille de voxel par sur-échantillonnage. Ou bien on peut se tourner vers d’autres procédures d’inférences (le *nonparametric framework*, l’approche FDR ou l’inférence bayésienne).

- Dans le *nonparametric framework* (voir chapitre 16 du livre), l'inférence sur les images statistiques avec des petits degrés de liberté peut être améliorée par l'usage de statistiques pseudo-t qui sont beaucoup plus lisses que les statistiques t correspondantes.
- La *False Discovery Rate* (FDR) contrôle la proportion de faux positifs parmi les voxels déclarés positifs (au lieu de l'erreur sur la famille et la probabilité d'avoir un faux positif dans tout le volume). Cette approche très différente est discutée au point 8.4.
- Finalement, le chapitre 17 du livre introduira l'inférence bayésienne où, au lieu de se focaliser sur l'improbabilité des données d'être issues d'une hypothèse nulle, les inférences sont fait sur la base d'une distribution postérieure qui caractérise notre paramètre d'incertitude sans faire référence à une distribution nulle.

8 La RFT, un pas plus loin

La RFT est utilisée dans l'analyse statistique de SPMs partout où il existe une composante spatiale à l'inférence. Le plus important est la **question de détecter un effet ou une activation à un endroit inconnu** (dans l'espace). Très souvent, nous ne savons pas à l'avance où chercher pour un effet. Nous sommes dès lors intéressés par une recherche dans l'ensemble du cerveau ou seulement une partie. Cela représente des problèmes statistiques spéciaux, liés au problème des comparaisons multiples ou tests multiples.

Deux méthodes ont été proposées pour résoudre ces problèmes, toutes deux utilisent la RFT :

1. une méthode basée sur le maximum de la statistique T ou F
2. une méthode basée sur l'étendue spatiale de la région où ces statistiques excèdent une valeur de seuil

8.1 Le maximum de la statistique de test

Une méthode simple pour trouver ces endroits est de choisir des endroits où un test statistique Z est grand (on pourrait très bien choisir aussi des statistiques sur T , χ^2 , F ou le T^2 de Hotelling). Pour ce faire, on applique un seuil de hauteur z à l'image de Z . Cela ne fait que repousser le problème : maintenant, le problème est de choisir ce seuil z de manière à exclure les faux positifs avec une grande probabilité (par exemple 0.95).

Si on définit z à la valeur critique de Z habituelle (non corrigée) $P = 0.05$ (1.64 dans le cas gaussien), cela signifie que 5% des parties inactivées du cerveau vont montrer de faux positifs. Nous avons besoin d'augmenter z de manière à ce que la probabilité de trouver une quelconque activation dans les régions activées soit de 0.05. C'est un type de comparaison multiple puisque nous sommes en train de tester l'hypothèse de l'absence d'activation sur un très grand nombre de voxels.

8.1.1 Application de la correction de Bonferroni

Une simple solution serait d'appliquer la correction de Bonferroni. La probabilité de détecter la moindre activation dans les régions inactivées est liée à la supposition que les régions inactivées couvrent l'entièreté de la région de recherche.

Suite à l'inégalité de Bonferroni (équation 2), la probabilité de détecter tout activation est liée à :

$$P(\max Z > z) \leq N \cdot P(Z > z) \quad (5)$$

Dans cette équation, le maximum est pris sur tous les N voxels présents dans la région de recherche.

Pour un test de statistique gaussienne $P = 0.05$, des seuils critiques de 4 ou 5 sont communs. Cependant, cette procédure est conservative si l'image est lisse (bien que les données de fMRI donnent souvent des seuils très précis).

8.1.2 Application de la RFT

La RFT donne une valeur de P moins conservative (plus petite) si l'image est lisse :

$$P(\max Z > z) \approx \sum_{d=0}^D \text{Resels}_d \cdot \text{EC}_d(z) \quad (6)$$

Dans cette équation,

- D est le nombre de dimensions de la région de recherche
- Resels_d est le nombre de resels de dimension d dans la région de recherche
- EC_d est la densité caractéristique d'Euler de dimension d

L'approximation 6 est basée sur le fait que le membre de gauche est l'espérance exacte de la caractéristique d'Euler pour la région au-dessus du seuil z . La caractéristique d'Euler compte le nombre d'ensembles (*clusters*) si la région n'a pas de trous, ce qui est très probablement le cas si z est grand.

8.1.3 Pratiquement ...

Pratiquement, l'approximation 6 est correcte pour des régions de recherche de toute taille ou forme, même un simple point ! Mais cette approximation est la meilleure pour des régions de recherche qui ne sont pas trop concaves. Parfois, il est même mieux d'entourer une région de recherche très convoluée (comme la matière grise) par une coque convexe (*convex hull*) avec un volume légèrement plus grand mais une surface plus petite ; cela permet d'avoir une valeur de P plus petite et plus précise.

Grandes régions de recherche Pour de larges régions de recherche, le dernier terme de la somme (membre droit de l'approximation 6, où $d = D$) est le plus important. Comme nous l'avons vu précédemment, le nombre de resels dépend de FWHM. Voici l'équation de cette dépendance :

$$\text{Resels}_D = \frac{V}{\text{FWHM}^D}$$

Dans cette équation,

- V est le volume de la région de recherche
- FWHM est la *Full Width at Half Maximum* effective d'un noyau gaussien utilisé pour lisser les données.

La densité EC correspondante pour une image statistique T avec ν degrés de liberté est (ν se prononce "nu") :

$$\text{EC}_3(z) = \frac{(4 \log_e 2)^{\frac{2}{3}}}{(2\pi)^2} \left(\frac{\nu - 1}{\nu} z^2 - 1 \right) \left(1 + \frac{z^2}{\nu} \right)^{-\frac{1}{2}(\nu - 1)}$$

Petites régions de recherche Pour de petites régions de recherche, les plus petits termes (de l'approximation 6) de dimensions $d < D$ deviennent importants. Cependant, la valeur de P n'est pas très sensible à la forme de la région de recherche. L'assomption d'une région de recherche sphérique donne une très bonne approximation.

La figure 14 montre le seuil z pour un test $P = 0.05$ calculé par les deux méthodes. Si le FWHM est petit par rapport à la taille du voxel, alors le seuil de Bonferroni est plus petit que

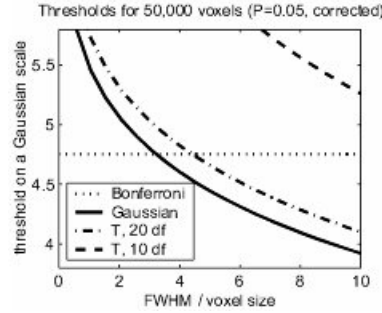


FIG. 14 – Seuils pour un volume de $N = 500000$ voxels ($P = 0.05$, corrigé). Notez que si la FWHM est plus petite que 3.2 voxels, la méthode de Bonferroni est meilleure que la méthode de la RFT pour une statistique gaussienne. Pour une statistique T avec $\nu = 20$ degrés de liberté, cette limite est plus grande (4.4) et encore plus grande (14.4, hors de l'échelle de cette figure) avec $\nu = 10$ degrés de liberté

celui de la RFT. En pratique, il est toujours mieux de prendre le plus petit des deux seuils (issus des équations 5 et 6).

Finalement, les densités d'EC pour d'autres tests peuvent être trouvés dans la littérature ainsi que les résultats pour les *correlation random fields*, les *space random fields*, etc. (voir le livre pour les références).

8.2 L'étendue spatiale maximale du test statistique

Une autre méthode simple pour chercher les endroits est basé sur l'étendue spatiale de clusters d'éléments connectés dont les voxels sont largement au-dessus du seuil : $Z > z$. z est choisi pour être approximativement 3 dans un champ de hasard gaussien. De nouveau, l'image doit être un champ de hasard stationnaire lisse.

L'idée est d'approcher la forme de l'image par une équation quadratique⁹ avec un pic au maximum local.

Pour un champ de hasard gaussien, l'étendue spatiale S est approchée alors par le volume où l'équation quadratique de la hauteur H au-dessus de z coupe le seuil z . Apparemment, cela s'écrit mathématiquement comme ceci :

$$S \approx cH^{\frac{D}{2}} \quad (7)$$

Dans cette équation,

$$c = \frac{\text{FWHM}^D \left(\frac{2\pi}{z}\right)^{\frac{D}{2}} (4 \log 2)^{-\frac{D}{2}}}{\tau \left(\frac{D}{2} + 1\right)}$$

Pour de grands z , la probabilité *upper tail* de H est bien approchée par :

$$P(H > h) = \frac{P(\max Z > z + h)}{P(\max Z > z)} \approx \exp(-zh) \quad (8)$$

⁹Une équation quadratique est une équation polynomiale de second ordre pour une seule variable X . Sa forme est $ax^2 + bx + c = 0$ (avec $a \neq 0$)

De cette dernière équation, nous pouvons conclure que H a une distribution approchée exponentielle avec une moyenne de $\frac{1}{z}$. A partir de là, nous pouvons trouver la valeur approchée de la valeur P de l'étendue spatiale S d'un seul cluster :

$$P(S > s) \approx \exp\left(-z \left(\frac{s}{c}\right)^{\frac{2}{D}}\right) \quad (9)$$

La valeur de P pour la plus grande étendue spatiale est obtenue simplement par la correction de Bonferroni pour le nombre attendu de clusters, K :

$$P(\max S > s) \approx E(K)P(S > s) \quad (10)$$

Avec $E(K) \approx P(\max Z > z)$ (l'équation 10 se trouve à partir de l'équation 6).

Nous pouvons améliorer de manière significative la valeur de la constante c en faisant correspondre l'étendue spatiale totale attendue, donnée par $V \cdot P(Z > z)$ à celle obtenue en additionnant les étendues spatiales de tous les clusters S_1, \dots, S_k :

$$VP(Z > z) = E(S_1 + \dots + S_k) = E(K)E(S)$$

En utilisant le fait que

$$E(S) \approx \frac{c\tau \left(\frac{D}{2} + 1\right)}{z^{\frac{D}{2}}}$$

de 7 et le nombre attendu de cluster de 6, il s'ensuit que :

$$c \approx \frac{FWHM^D z^{\frac{D}{2}} P(Z > z)}{EC_D(z)\tau \left(\frac{D}{2} + 1\right)}$$

Le livre cite des extensions de ces résultats à d'autres statistiques.

8.2.1 La recherche dans de petites régions

Pour de petites régions de recherche pré-définies comme le cingulate, les valeurs de P pour le maximum de la statistique de test sont très bien estimée par l'équation 6 mais les résultats de cette section (étendue spatiale maximale) ne s'appliquent qu'à de grandes régions de recherche.

Friston a proposé une méthode (apparemment fascinante : c'est le livre qui le dit) qui évite le problème maladroit de la spécification préalable de la petite région de recherche.

Nous appliquons un filtre z à l'image de test statistique. Ensuite, nous prenons simplement le pic le plus proche d'un point ou d'une région d'intérêt. La partie intelligente est la suivante ...

Puisque nous avons identifié ce pic seulement sur sa localisation spatiale et non sur sa hauteur ou son étendue, il n'y a maintenant pas besoin de corriger quoi que ce soit pour chercher sur tous les pics.

Donc, la valeur P pour son étendue spatiale S est simplement $P(S > s)$ de l'équation 9. Et la valeur de P pour sa hauteur de pic H au-dessus de z est simplement $P(H > h)$ de l'équation 8.

8.3 Quelques détails supplémentaires

8.3.1 L'estimation de la FWHM

Le seul composant dépendant des données requis pour définir les seuils évoqués ci-dessus est $Resels_D$ et, indirectement, la FWHM.

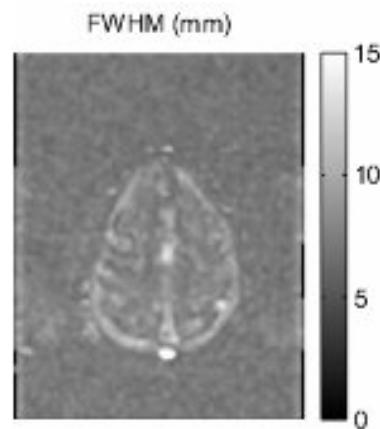


FIG. 15 – La FWHM estimée pour une tranche de données de fMRI (image reprise du livre)

La FWHM dépend souvent de la localisation : les données de fMRI sont considérablement plus lisses dans le cortex que dans la matière blanche (voir figure 15) et les données de VBM [?], la FWHM varie considérablement d'un lieu à l'autre. **Cela signifie que le champ de hasard n'est pas isotrope et donc que la théorie ci-dessus n'est pas valide.** Heureusement, il y a une manière simple de permettre cela en estimant la FWHM pour chaque voxel.

Développement mathématique dans le livre ...

8.4 La False Discovery Rate

9 Et où allons-nous, maintenant ?

Dans ce chapitre, nous nous sommes penché sur l'inférence au niveau du voxel basée sur les seuils de hauteur. La question que nous avons essayé de résoudre était : "est-ce que l'activation à un voxel donné est significativement non nulle?". De manière plus globale, l'inférence au niveau d'un voxel peut être replacée dans un cadre plus général incluant les inférences aux niveaux d'ensemble et de groupes.

Ces autres inférences demandent que l'utilisateur fixe les seuils de hauteur et d'extension spatiale (*spatial extend*). Les valeurs de p corrigées (*corrected p-values*) peuvent alors être dérivée en ce qui concerne :

1. le nombre de régions activées (càd. le nombre de clusters au-dessus des seuils de hauteur et de volume) – *set level inferences*
2. le nombre de voxels activés (càd. le volume) dans une certaine région – *cluster level inferences*
3. la valeur p (*p-value*) de chaque voxel à l'intérieur de chaque cluster – *voxel level inferences*

Ces valeurs de p sont corrigées pour les comparaisons multiples dépendantes et sont basées sur la probabilité d'obtention de c (ou plus) clusters avec k (ou plus) voxels au-dessus d'un seuil u dans une SPM de caractère lisse connu ou estimé.

Les *set level inferences* peuvent être plus puissantes que les *cluster level inferences* qui, elles-mêmes sont plus puissantes que les *voxel level inferences*. Le prix à payer pour cette augmentation de sensibilité est un pouvoir de localisation diminué.

Sensibilité : voxel level inferences < cluster level inferences < set level inferences

Pouvoir de localisation : set level inferences < cluste level inferences < voxel level inferences

Les tests au niveau du voxel permettent la caractérisation de voxels individuels comme significatifs alors que les *cluster-* et *set-level inferences* ne permettent qu'à des clusters ou sets d'être significatifs.

On doit cependant se rappeler que ces conclusions (sur les pouvoirs relatifs des différentes inférences) sont basées sur des activations distribuées. Une activation focale peut très bien être détectée avec une plus grande sensibilité par des tests au niveau du voxel, basés sur la hauteur du pic! De manière typique, les gens utilisent des inférences au niveau du voxel et un seuil d'étendue spatiale de zéro. Cela reflète le fait que les caractérisations d'anatomie fonctionnelle sont généralement plus utiles lorsqu'elle sont spécifiées avec un grand degré de précision anatomique.

La section suivante s'appelle l'intégration fonctionnelle [creuser un peu].

Références

- [1] R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, C.J. Price, S. Zeki, J. Ashburner, and W.D. Penny. *Human Brain Function*. Academic Press, 2nd edition, 2003.
- [2] K.J. Worsley, S. Marrett, P. Neelin, and A.C. Evans. A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12 :900–918, 1992.