All-Resolutions Inference for brain imaging

- Article étudié : All-Resolutions Inference for brain imaging de Rosenblatt, Finos, Weeda, Solari et Goeman
- Publié en 2018 dans la revue NeuroImage
- Etude de l'activation des zones du cerveau pour différentes tâches
- Objet de l'article : nouvelle méthode pour l'inférence statistique en imagerie cérébrale

Cartographie du cerveau humain : un problème important et toujours d'actualité

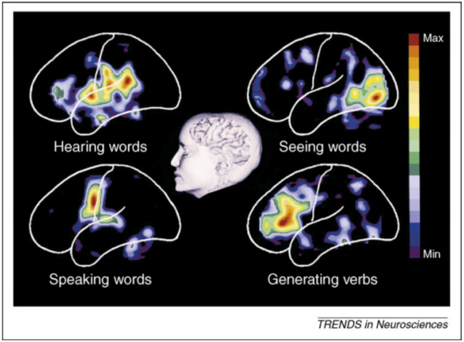


Figure – Un exemple d'imagerie cérébrale dans l'étude du langage (1988)

Idée : on dispose d'un IRM "au repos" et d'un IRM "exécutant une action"

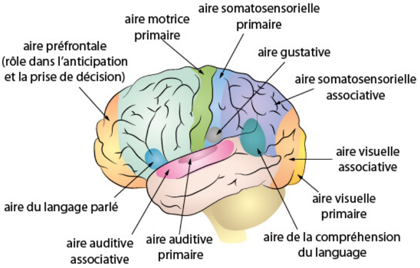


Figure – Un exemple de cartographie du cerveau

$\hookrightarrow$ Pour chaque voxel i : test d'hypothèses $H_{0,i}$: "voxel inactif" contre $H_{1,i}$: "voxel actif"

- Une observation apparemment statistiquement significative peut en fait surgir par hasard en raison de la taille même de l'espace des paramètres à rechercher.

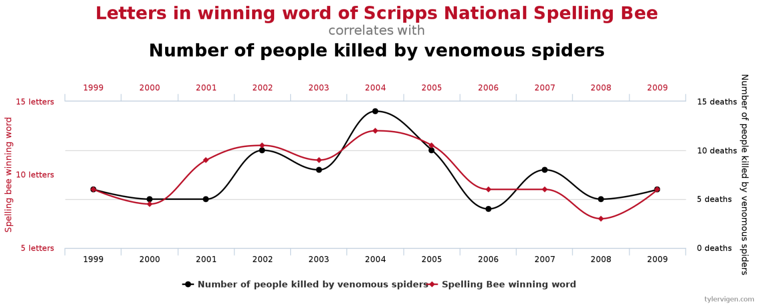


Figure – Un exemple de corrélation fallacieuse

- Dans une image de 100 000 voxels, même dans le cas où aucun voxel est activé : 5000 faux positifs attendus.

Le contrôle de l'erreur de type 1 pour les tests multiples

FWER (Family Wise Error Rate), la probabilité de rejeter au moins un test à tort, contrôlable par la **correction de Bonferroni**

Principe : on considère les $|C|$ p-valeurs p_i associées au cluster C

On rejette H_0 au niveau α si une des p-valeurs est inférieure à $\alpha/|C|$

$\hookrightarrow$ Autre approche classique : contrôle du FDR (False Discovery Rate), la proportion moyenne de faux positifs, par la procédure de **Benjamini-Hochberg**

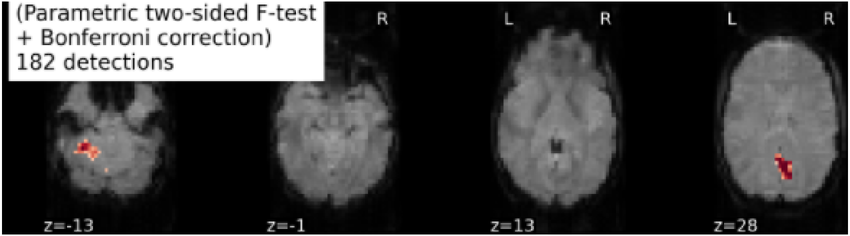


Figure – Inférence par correction de Bonferroni

Contrôle de l'erreur de type I mais faible puissance statistique

Interdiction de choisir un nouveau cluster plus ciblée -> Problème de l'inférence sélective

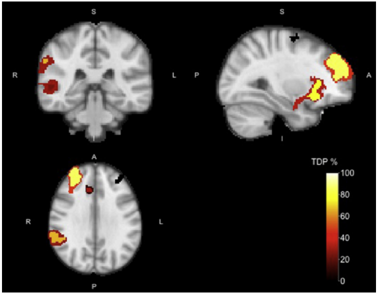


Figure – Un exemple d'inférence tiré de l'article de Rosenblatt et al., 2018

Algorithme :

- Seuillage des p-valeurs -> premiers clusters
- Calcul borne ARI dans ces clusters (légitime car borne post-hoc)
- Nouveau seuillage plus strict -> nouveaux clusters
- Calcul borne ARI dans ces clusters (légitime car borne post-hoc)

On note :

- m = nombre d'hypothèses
- S = sous-ensemble d'hypothèses d'intérêt
- $a(S)$ = Proportion de vraies détections dans S (TDP)

$\hookrightarrow$ Borne inférieure du nombre de vrais positifs au niveau de risque α est une fonction V(S) telle que :

$$\mathbb{P}\left(a(S) \geq V(S)\right) \geq 1 - \alpha$$

$\hat{S}$ = sous-ensemble de S choisi en fonction des données
Exemple typique = plus petites p-valeurs

$\hookrightarrow$ On s'attend à avoir l'inégalité :

$$\mathbb{P}\left(a(\hat{S}) \geq V(\hat{S})\right) \geq 1 - \alpha$$

Or ceci est faux en général ! Pourquoi ?

$\hookrightarrow$ Dans l'inégalité écrite plus haut, il y a deux sources d'aléatoire :
plus la même loi de probabilité !

Voir Kriegeskorte et al., *Circular analysis in systems neuroscience : the dangers of double dipping*, Nature 2009

Solution possible au problème d'inférence sélective : fonction V(S) telle que

$$\mathbb{P}\left(\forall S, a(S) \geq V(S)\right) \geq 1 - \alpha$$

$\hookrightarrow$ **Borne post-hoc**, ici on peut sélectionner des sous-ensembles d'hypothèses en utilisant les données

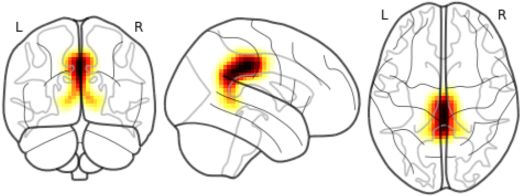
Borne ARI

$$V(S) = \min\left\{0 \leq k \leq |S| : \min_{1 \leq i \leq |S|-k} \frac{h(\alpha)}{i} p_{(i+k:S)} > \alpha\right\}$$

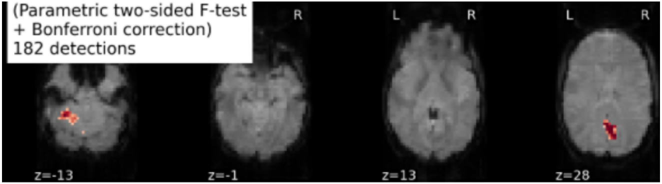
Avec

$$h(\alpha) = \max\left\{i \in \{0, \ldots, m\} : i p_{(m-i+j:B)} > j\alpha, \quad \text{for } j = 1, \ldots, i\right\}$$

Exemples d'usage du package Nilearn :



(a) Visualisation avec Nilearn



(b) Inférence avec Nilearn

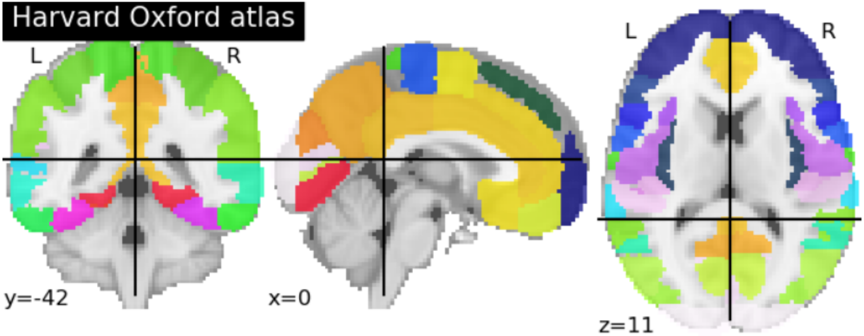


Figure – Atlas Harvard-Oxford sur Nilearn

Résultats : ARI vs Bonferroni vs maxT

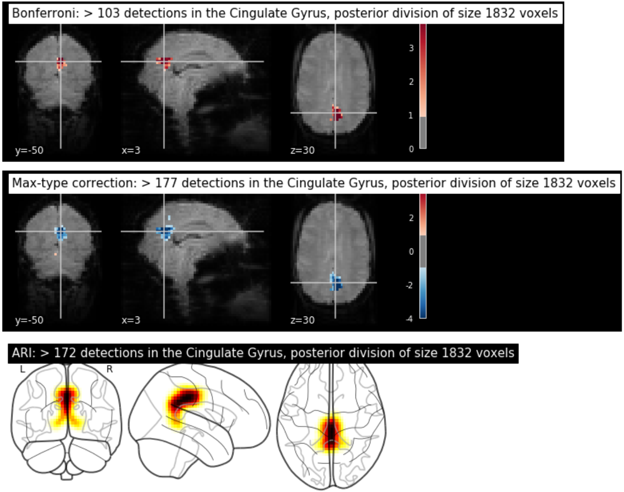


Figure – Résultats sur le Gyrus Cingulaire

Nous avons distingué trois types de limites :

- **Limite des hypothèses et des a priori de la modélisation**
- Limite de l'applicabilité
- Critique de la validation des résultats

- Présupposé dans l'article : on dispose de p-valeurs
- Pas de contrôle de l'erreur de type 2
- Activation binaire ?

- Résolution du problème de multiplicité et de circularité
- Borne post-hoc qui donne des résultats proches de la méthode max-type conseillée dans Nilearn
- Mais validation discutable