Tests de permutation

Abdoulaye Diabakhaté
Encadrants: Stéphane ROBIN et Mahendra MARIADASSOU

23 octobre 2018

Plan de la présentation

Définition

Les approches paramétriques standard exigent plusieurs suppositions d'application concernant le plan de l'expérience (échantillonnage aléatoire) et le modèle de population (distribution normale ou homoscédasticité, par exemple).Lorsque les conditions d'application de ces approches ne sont pas respectées, en particulier lorsque la loi des données n'est pas conforme aux exigences du test, les résultats des tests paramétriques sont moins fiables.

Les tests non paramétriques offrent une alternative importante puisqu'ils nécessitent moins de suppositions.
L'importance des tests de permutation réside dans leur flexibilité et leur robustesse lorsque les suppositions statistiques des tests paramétriques habituels ne sont pas valides.
Pour appliquer les tests de permutation,il faut que la condition nécessaire et suffisante sous l'hypothèse nulle,l'échangeabilité des observations,soit vérifiée.

Test de permutation (ou test par randomisation) :

- Ne pas confondre avec non paramétrique
- Classe de test qui se différencie par la façon de calculer la densité de probabilité d'une statistique de test
- Souvent utilisés pour les tests non paramétriques, mais valable aussi en paramétrique

échangeabilité

Soit Π,l'ensemble des permutations des entiers (1,...,n).Un ensemble de variables aléatoires $X_1, ..., X_n$ est échangeable, si pour toute permutation $\pi = (\pi_1, ..., \pi_n) \in \Pi$, on a :

- $(X_1, ..., X_n) \underset{d}{=} (X_{\pi_1}, ..., X_{\pi_n})$

Où d, indique que les deux vecteurs ont la même fonction de répartition (égalité en loi).

L'exemple le plus simple de variables échangeable est le cas où $X_1, ..., X_n$ sont des variables indépendantes et identiquement distribuées.Dans ce cas, les variables sont échangeable car :

$f(x) = f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i) = f(x_{\pi_1}, ..., x_{\pi_n})$

Construction des distributions sous H_0 et calcul des p-values

- On calcule classiquement la statistique de test pour les k gènes : $t_1, t_2, ..., t_k$

- On réalise B itérations. Pour chaque itération (b=1,2,...,B) :

1- Permutation des colonnes.

2- Calcul des statistiques de test pour chaque gène : $t_1^{(b)}, t_2^{(b)}, ..., t_k^{(b)}$

- Calcul des probabilités critiques :
  $p_j^* = \frac{\sum_{b=1}^{B} I(|t_j^{(b)}| \geq |t_j|)}{B}$

  Où I(a) est égale 1 si la condition a est respectée et 0, sinon.

Remarque

Les permutations ont pour conséquence de casser les relations entre la variable dépendante et la ou les variables indépendantes. Lorsque la relation est détruite, on est dans la situation de l'hypothèse nulle, qui postule que le groupe n'a pas d'effet.
Les tests de permutation permettent donc de construire la distribution empirique de la statistique sous l'hypothèse nulle.