



Classification supervisée par modèle de mélange: application aux diagnostics par autopsie verbale

Seydou N. SYLLA^{1,2,3}, Dr Aldiouma DIALLO¹, Dr Cheikh SOKHNA¹,
Pr Abdou Ka DIONGUE², Pr Stéphane GIRARD³

¹URMITE-IRD-Dakar, ²LERSTAD-UGB-Saint-Louis, ³Mistis-INRIA-Grenoble

Données

Les données analysées portent sur les symptômes et les causes probables de décès concernant des personnes décédées de 1985 à 2010 dans les trois sites de l'IRD (Niakhar, Bandafassi et Mlomp) au Sénégal.

Les variables aléatoires X représentent les symptômes.

Les variables aléatoires Z représentent les sous-items pour chaque variable X .

La variable aléatoire C représente le groupe à priori (diagnostics des médecins).

L'analyse porte sur $n = 2500$ individus répartis sur $K = 22$ classes et caractérisés par plus de 30 variables.

$$p(C|x) = \frac{p(x|C)p(C)}{p(x)}$$

$$g(x, \Phi) = \sum \pi_k f(x, \theta_k)$$

Modèles de mélange hiérarchique

✓ Un modèle de mélange sous l'hypothèse d'indépendance conditionnelle concernant seulement les variables aléatoires X représentant les données symptomatiques et asymptotiques (âge, saison, sexe, ...).

✓ Un modèle de mélange hiérarchique qui prend en compte les variables aléatoires Z représentant les sous-items de chaque variable X .

Dans ces modèles, les classes sont les différents diagnostics établis par les médecins après l'analyse de l'autopsie verbale.

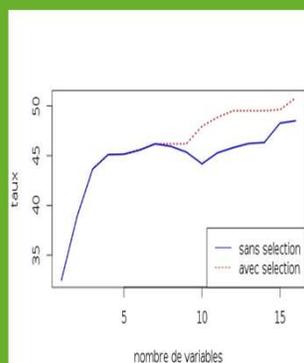


Figure2: Sélection de variables: méthode stepwise

Résultats

❖ Les résultats préliminaires montrent 50 % de concordance sur 20 causes de décès identifiées par les médecins après diagnostics par autopsies verbales.

❖ Les causes de décès regroupées en 6 classes montrent un taux de concordance de 60% .

Commentaires

Le modèle développé dépend de l'homogénéité des classes.

Vu la complexité du diagnostic par autopsie verbale, les variables les plus pertinentes ont été incluses dans le modèle.

Ce modèle pourrait être une alternative pour le diagnostic des causes de décès établies par autopsies verbales dans les sites de suivis démographiques.

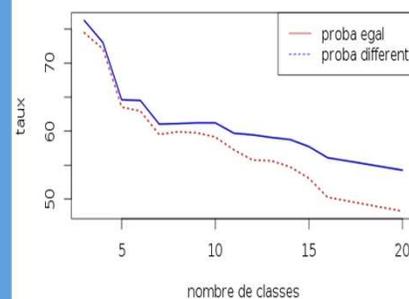


Figure1 : taux de concordance en fonction du nombre de causes