
Package BlockSeg pour la détection rapide des frontières des blocs d'une matrice constante par blocs bruitée

Vincent Brault^{*1,2}, Julien Chiquet¹, and Céline Lévy-Leduc¹

¹AgroParisTech (AgroParisTech) – Institut national de la recherche agronomique (INRA) : UMR518, AgroParisTech – AgroParisTech 75231 Cedex 05, Paris - France, France

²INRA – Institut national de la recherche agronomique (INRA), Institut National de la Recherche Agronomique - INRA – France

Résumé

Dans différents contextes, on peut être amené à partitionner les lignes et les colonnes d'une matrice pour former un quadrillage de blocs homogènes sans effectuer de permutations ; c'est notamment le cas pour l'analyse des données Hi-C qui mesurent le degré d'interaction physique entre différentes positions du génome (Dixon et al.). En effet, ces données peuvent être modélisées comme une matrice constante par blocs bruitée. Toutefois, ce problème peut être compliqué pour plusieurs raisons : les méthodes utilisées en segmentation unidimensionnelle comme l'algorithme de programmation dynamique ne s'appliquent pas dans ce cas et la taille importante des données nécessite le développement et la mise en place d'algorithmes performants.

Notre approche est la suivante : nous montrons que ce problème peut être ramené à celui d'un modèle linéaire parcimonieux de grande dimension pour lequel nous proposons une méthode de sélection de variables rapide et efficace.

Dans cet exposé, nous présenterons le package **blockseg** où sont implémentées les méthodes. Après avoir rappelé le modèle et les estimations, nous illustrerons le comportement de l'algorithme à l'aide de film puis nous expliciterons les différents programmes et sorties graphiques associées.

*Intervenant