



HAL
open science

Clustering et sélection de variables sur des données génétiques

Dominique Bontemps, Wilson Toussile

► **To cite this version:**

Dominique Bontemps, Wilson Toussile. Clustering et sélection de variables sur des données génétiques. Journées MAS et Journée en l'honneur de Jacques Neveu, Aug 2010, Talence, France. inria-00510368

HAL Id: inria-00510368

<https://hal.inria.fr/inria-00510368>

Submitted on 18 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journées MAS 2010, Bordeaux

Session : Sélection de modèles

Clustering et sélection de variables sur des données génétiques

par **Dominique Bontemps** et Wilson Toussile

Nous nous intéressons au problème d'estimer les variables pertinentes et le nombre de composantes d'une loi de mélange pour des données génotypiques multilocus. Un critère du maximum de vraisemblance pénalisé est proposé, et une inégalité oracle non-asymptotique est obtenue. En outre, sous des conditions faibles portant sur la distribution qui a généré les observations, le modèle sélectionné est asymptotiquement consistant. D'un point de vue pratique, la pénalité est définie à une constante multiplicative près, et celle-ci est calibrée par l'heuristique de pente. Sur des données simulées la procédure de sélection fait mieux que des critères classiques tels que BIC et AIC. Le nouveau critère apporte une réponse à la question : "Quel critère choisir en fonction de la taille de l'échantillon ?".

Adresses :

Dominique BONTEMPS
Département de Mathématiques
Univ. Paris-Sud 11
Bât. 430
91405 Orsay France
E-mail : dominique.bontemps@math.u-psud.fr
<<http://www.math.u-psud.fr/~bontemps/>>

Wilson TOUSSILE
Département de Mathématiques
Univ. Paris-Sud 11
Bât. 430
91405 Orsay France
E-mail : Wilson.Toussile@math.u-psud.fr
<<http://www.math.u-psud.fr/~toussile/>>

Session : Sélection de modèles