

# CENTRE D'ETUDES DOCTORALES «SCIENCES ET TECHNIQUES ET SCIENCES MÉDICALES »

مركز الدكتوراء « الطبية» والتقنيات على الطبية المالية المالية المالية المالية المالية المالية المالية المالية ا

# AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz -Fès - annonce que

# Mr ALOUI Abdelaziz Soutiendra : le Samedi 01/11/2025 à 10H00

Lieu: Centre des Etudes Doctorales - USMBA - Amphi 2

## Une thèse intitulée:

# «Sélection de modèle de mélange par les critères de l'information»

En vue d'obtenir le **Doctorat** 

FD : Sciences et Technologies de l'Information et de la Communication Spécialité : Probabilités et Statistique

## Devant le jury composé comme suit :

Nom et prénom	Etablissement	Grade	Qualité
OUASSOU Idir	École Nationale des Sciences Appliquées, Marrakech	PES	Président
KIOUACH Driss	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Rapporteur
BENMOUMEN Mohammed	Faculté des Sciences, Oujda	PES	Rapporteur
EL MELHAOUI Said	Faculté des Sciences Juridiques, Économiques et Sociales, Oujda	PES	Rapporteur
RACHDI Mustapha	Université Pierre-Mendès-France, Grenoble	PES	Examinateur
RHOMARI Noureddine	Faculté des Sciences, Oujda	PES	Examinateur
EL MATOUAT Abdelaziz	Ecole Normale Supérieure, Fès	PES	Invité
HAMZAOUI Hassania	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Directeur de thèse



CENTRE D'ETUDES DOCTORALES «SCIENCES ET TECHNIQUES ET SCIENCES MÉDICALES »

مركز الدكتوراء « الطبية» والتقنيات

#### Résumé:

Le sujet de cette thèse est d'étudier le problème de sélection de modèles de mélange par les critères d'information. Nous nous intéressons principalement aux critères d'Akaike AIC et de Schwarz BIC. Le critère de Schwarz BIC est convergent mais il peut conduire à une sousparamétrisation du modèle lorsque la taille de l'échantillon est petite. Nous avons élaboré un critère de Schwarz modifié, noté BICr, en utilisant une technique de rééchantillonnage pour la sélection d'un modèle paramétrique quelconque tout en préservant sa propriété de convergence. La performance du nouveau critère par rapport aux critères AIC et BIC a été montrée sur des données simulées de modèles de mélange gaussien. Le critère d'Akaike AIC n'est pas convergent, il conduit asymptotiquement à une surpramétrisation stricte du modèle, et lorsque la taille de l'échantillon est petite, il peut conduire à une surestimation du nombre de composantes et du nombre de variables explicatives pour les modèles de mélange de régressions. Nous avons dérivé un nouveau critère, noté MRCv, pour l'estimation conjointe du nombre de composantes et du nombre de variables explicatives pour les modèles de mélange de régressions multivariées à partir de la log-vraisemblance des données complètes et de l'information de Kullback-Leibler. Avec des données simulées de modèles de mélange de régressions multivariées, nous avons montré l'intérêt du critère MRCv par rapport aux critères AIC et BIC. Ensuite, dans le même contexte, nous avons proposé un critère d'information, noté MRCsdv, basé sur la divergence symétrique de Kullback-Leibler et la log-vraisemblance des données complètes. Nous avons mis en évidence les performances du critère MRCsdv par rapport au critère MRCv sur des données simulées de modèles de mélange de régressions multivariées.

En présence de données manquantes, nous avons considéré la probabilité a posteriori des modèles candidats et l'algorithme EM pour développer un critère de type BIC, noté BICcd. La performance de ce critère par rapport au critère AICcd élaboré par Cavanaugh et al. et les critères AIC et BIC a été illustrée sur des données simulées de modèles autorégressifs et de modèles de mélange de régressions univariées.

### Mots clés :

Sélection de modèles, Critères d'information, Algorithme EM, Modèles de mélange de lois gaussiennes, Modèles de mélange de régressions

# MIXTURE MODELS SELECTION BY INFORMATION CRITERIA

#### **Abstract:**

The subject of this thesis is the study of the problem of a mixture model selection using information criteria. We are mainly interested in the Akaike criterion AIC and the Schwarz



CENTRE D'ETUDES DOCTORALES «SCIENCES ET TECHNIQUES ET SCIENCES MÉDICALES »

مركز الدكتوراء « الطبية» هايقنبإيت عنوالية الطبية الطبية

criterion BIC. The Schwarz criterion BIC is convergent but it can lead to a subparametrization of the model when the sample size is small. We developed a modified Schwarz criterion, denoted BICr, using a resampling technique for the selection of any parametric model while preserving its convergence property. The performance of the new criterion compared to the criteria AIC and BIC was shown on simulated data from gaussian mixture models. The Akaike criterion AIC is not convergent, it leads asymptotically to a strict overprametrization of the model, and when the sample size is small, it can lead to an overestimation of the components number and the explanatory variables number for mixture regression models. We derived a new criterion, denoted MRCv, for the joint estimation of the components number and the explanatory variables number for mixture multivariate regression models using the complete-data log-likelihood and the Kullback-Leibler information. With simulated data from a mixture multivariate regression model, we showed that the criterion MRCv outperforms the criteria AIC and BIC. In the same context, we proposed an information criterion, denoted MRCsdy, based on the Kullback-Leibler symmetric divergence and the complete-data log-likelihood. We highlighted the performance of the criterion MRCsdv compared to the criterion MRCv on simulated data from mixture multivariate regression models. In the presence of missing data, we considered the a posteriori probability of candidate models and the EM algorithm to develop a criterion of type BIC, denoted BICcd. The performance of this criterion compared to the criterion AICcd developed by Cavanaugh et al. and the criteria AIC and BIC was illustrated on simulated data from autoregressive models and univariate regression mixture models.

## **Key Words:**

Model selection, Information criteria, Resampling technique, EM algorithm, Gaussian mixture models, Regression mixture models