# Sélection de modèles de mélange de régression multivariée

Abdelaziz Aloui [1], Abdelaziz El Matouat [2] & Hassania Hamzaoui [3]

[1] *FSDM Université de Fès, Maroc, abdelaziz.aloui@usmba.ac.ma*
[2] *LMAH Université du Havre, France,abdelaziz.el-matouat@univ-lehavre.fr*
[3] *FLSH Université de Fès, Maroc, hassania.hamzaoui@usmba.ac.ma*

**Résumé.** Dans la détermination du nombre de composantes et l'ordre d'un modèle de mélange de régression univarié, Naik et al(2007) ont développé le critère $MRC$ (Mixture Regression Criterion). Dans ce travail, nous proposons une généralisation de ce critère au modèle de mélange de régression multivarié et à un modèle de mélange autorégressif vectoriel.

**Mots-clés.** Critères d'information, critère $MRC$, sélection de modèle, régression multivariée, modèle autorégressif vectoriel.

**Abstract.** In determining the number of components and the order of the mixture of univaraite regression model, Naik and al(2007) have developed the $MRC$ criterion (Mixture Regression Criterion). In this work, we propose a generalisation of this criterion to the mixture multivariate regression model and mixture vector autoregressive model.

**Keywords.** Information criteria, MRC criterion, model selection, multivariate regression, vector autoregressive model.

# 1 Introduction

Un modèle de mélange linéaire multivarié consiste à écrire un vecteur $y$ en fonction des covariables $x$ de la façon suivante :

$$y = \begin{cases} \beta_1' x + \epsilon_1 & \text{avec la probabilité } \alpha_1 \\ \beta_2' x + \epsilon_2 & \text{avec la probabilité } \alpha_2 \\ . \\ . \\ \beta_K' x + \epsilon_K & \text{avec la probabilité } \alpha_K \end{cases}$$

où $y = (y_1, ..., y_m)'$ , $x = (x_1, ..., x_p)'$, $\beta_k$ une matrice d'ordre $p \times m$, et $\epsilon_k$ un vecteur aléatoire représentant les erruers de spécification du modèle, de dimension $m$ de loi gaussienne de moyenne 0 et de matrice de variances-covariances $\Sigma_k$.

La densité de probabilité relative à la loi de $y$ conditionnellement à $x$ est donnée par :

$$f(y,x,\phi) = \sum_{k=1}^{K} \alpha_k f_k(y,x,\beta_k,\Sigma_k) \tag{1}$$

où $\phi = \{(\alpha_k, \beta_k, \Sigma_k), k = 1, ....., K\}$ est l'ensemble des paramètres tel que $0 < \alpha_k < 1$ et $\Sigma_{k=1}^{K} \alpha_k = 1$, $K$ est le nombre de composantes et $f_k$ est la fonction densité d'une loi normale $N(\beta_k' x, \Sigma_k)$.
Soit $\{(y_1, x_1), ..., (y_n, x_n)\}$ un échantillon observé issu du modèle (1), la fonction log-vraisemblance s'écrit :

$$\begin{array}{rcl} L(\phi, Z, Y, X) & = & \log \prod_{j=1}^{n} \prod_{k=1}^{K} \{\alpha_k f_k(y_j, x_j, \beta_k, \Sigma_k)\}^{z_{jk}} \\ & = & \sum_{j=1}^{n} \sum_{k=1}^{K} z_{jk} \{log \alpha_k + log f_k(y_j, x_j, \beta_k, \Sigma_k)\} \end{array} \tag{2}$$

où $Y = (y_1, ..., y_n)'$, $Z$ est une variable latente d'ordre $n \times K$ telle que $z_{jk}$ vaut 1 lorsque $y_j$ provient de la $\mathrm{k}^{\mathrm{ième}}$ composante du mélange et à 0 sinon, et $X = (x_1, ..., x_n)'$ est la variable explicative.
On note $f^0$ la densité du vrai modèle telle que :

$$f^0(y,x,\phi^0) = \sum_{k=1}^{K^0} \alpha_k^0 f_k^0(y,x,\beta_k^0,\Sigma_k^0)$$

$L^0$ est la log-vraisemblance avec $1 < K^0 < K$ et $1 < p^0 < p$. Sous cette hypothèse, les colonnes de $X$ peuvent être réorganisées afin que $X^0 \beta_k^0 = X \beta_k^*$ avec $\beta_k^* = ((\beta_k^0)', (\beta_k^1)')'$ et $\beta_k^1$ est une matrice nulle d'ordre $(p - p^0) \times m$ [3].

# 2 Critère MRC pour un modèle de mélange de régression multivarié

## 2.1 Estimation des paramètres

Pour estimer les paramètres d'un modèle de mélange de régression, on applique l'algorithme EM (Expectation-Maximization)[2]. Cet algorithme est une méthode d'estimation itérative composée à l'itération $q$ de deux étapes :
**Etape E (Expectation) :** on calcule l'espérance $Q(\phi, \phi^{(q)})$ de la vraisemblance $L$ conditionnellement aux données observées qui nécessite principalement le calcul des probabilités a posteriori $\tau_{jk}^{(q)} = E(z_{jk}/y_j) = P(z_{jk} = 1/y_j\ )$. Par la formule de Bays on a :

$$\tau_{jk}^{(q)} = \frac{\alpha_k^{(q)} f_k(y_j, x_j, \beta_k^{(q)}, \Sigma_k^{(q)})}{\sum_{k=1}^{K} \alpha_k^{(q)} f_k(y_j, x_j, \beta_k^{(q)}, \Sigma_k^{(q)})}$$

**Etape M (Maximization) :** on calcule le vecteur paramètre $\phi^{(q+1)}$ qui maximise, par rapport à $\phi$, la fonction $Q$ qui s'écrit comme suit :

$$Q(\phi,\phi^{(q)}) = \sum_{j=1}^{n}\sum_{k=1}^{K}\tau_{jk}^{(q)}\{log\alpha_k + log f_k(y_j,x_j,\beta_k,\Sigma_k)\}$$

Les estimateurs des paramètres sont :

1. $\alpha_k^{(q+1)} = \sum_{j=1}^{n}\frac{\tau_{jk}^{(q)}}{n}$
2. $\beta_k^{(q+1)} = (\tilde{X}_k^{(q)\prime}\tilde{X}_k^{(q)})^{-1}\tilde{X}_k^{(q)\prime}\tilde{Y}_k^{(q)}$
3. $\Sigma_k^{(q+1)} = \tilde{Y}_k^{(q)\prime}(I-\tilde{H}_k^{(q)})\tilde{Y}_k^{(q)}/tr(W_k^{(q)})$

Où $W_k^{(q)} = diag(\tau_k^{(q)})$, $\tau_k^{(q)} = (\tau_{1k}^{(q)},...,\tau_{nk}^{(q)})'$
$\tilde{Y}_k^{(q)} = (W_k^{(q)})^{1/2}Y$, $\tilde{X}_k^{(q)} = (W_k^{(q)})^{1/2}X$ et $\tilde{H}_k^{(q)} = \tilde{X}_k^{(q)}(\tilde{X}_k^{(q)\prime}\tilde{X}_k^{(q)})^{-1}\tilde{X}_k^{(q)\prime}$. Les vecteurs paramètres $\phi^{(q)}$ convergent vers l'estimteur du maximum de vraisemblance $\hat{\phi}$.

## 2.2 Estimation du nombre de composantes et de l'ordre du modéle de mélange de régression multivariée :

L'objectif principal de la sélection du modèle est d'approcher le vrai modèle avec des modèles candidats en faisant varier le couple $(K,p)$, cette approche entraîne une conservation du modèle qui entraîne une perte minimale de l'information. Pour cette raison, nous adoptons l'approche prédictive d'Akaike(1985) pour déduire une nouvelle formulation du critère $MRC$. On considère :

$$I(\theta^0,\theta) = E_{Y^*/\theta^0}\{L^0(\phi^0,Z^*,Y^*,X) - L(\phi,\tau,Y^*,X)\}, \qquad \theta=(\phi,\tau) \tag{3}$$

$Y^* = (y_1^*,...,y_n^*)'$ est la prédiction, $Z^*$ est une matrice d'ordre $n\times K^0$, $\tau$ est une matrice telle que $\tau_{jk} = E(z_{jk}/y_j)$, et $E_{Y^*/\theta^0}$ est l'espérance conditionnelle sachant $\theta^0 = (\phi^0,\tau^0)$. En omettant le premier terme de (3) qui ne dépend pas de $\theta$, nous avons la proposition suivante :
**Proposition :** Sous des conditions de régularité le critère défini par :

$$MRC = \sum_{k=1}^{K}\hat{n}_k log(|\hat{\Sigma}_k|) - 2\sum_{k=1}^{K}\hat{n}_k log(\hat{\alpha}_k) + \sum_{k=1}^{K}\hat{d}_k m(p_k+\hat{n}_k) \tag{4}$$

est un estimateur de $\Delta(\hat{\theta}(Y)) = -2E_Y[E_{Y^*}\{L(\hat{\phi}(Y),\hat{\tau}(Y),Y^*,X)\}]$
avec $\hat{\theta} = (\hat{\phi},\hat{\tau})$, $\hat{n}_k = tr(\hat{W}_k)$, $\hat{d}_k = \frac{\hat{n}_k}{\hat{n}_k-(m+p_k+1)}$ et $p_k = tr(\hat{H}_k)$
**Preuve : (Idée de la démonstration)**

1. Soit $\Delta(\hat{\theta}(Y)) = -2E_YE_{Y^*}[\sum_{k=1}^{K}\sum_{j=1}^{n}\hat{\tau}_{jk}\{\log\hat{\alpha}_k + \log f_k(y_j^*,x_j,\hat{\beta}_k,\hat{\Sigma}_k)\}]$ (1-1)

2. Après des simplifications algébriques nous abtenons :
$\Delta(\hat{\theta}(Y)) = E_Y\{\sum_{k=1}^{K^0} tr(\hat{W}_k)\log(|\hat{\Sigma}_k|) - 2\sum_{k=1}^{K^0} tr(\hat{W}_k)`\log(\hat{\alpha}_k)\}$
$+E_Y\{\sum_{k=1}^{K^0} tr[\hat{X}_k(\beta_k^* - \hat{\beta}_k)\hat{\Sigma}_k^{-1}(\beta_k^* - \hat{\beta}_k)'\hat{X}_k']\} + E_Y\{\sum_{k=1}^{K^0} tr(\hat{W}_k)tr(\hat{\Sigma}_k^{-1}\Sigma_k^0)\}$ (1-2)
3. D'après les conditions de régularité, $\hat{\tau}_k$ est un estimateur consistant de $\tau_k^0 = (\tau_{1k}^0, ..., \tau_{nk}^0)$ [4] où $\tau_{jk}^0 = E(z_{jk}^0/Y)$. Par conséquent, nous pouvons remplacer $\hat{W}_k$ et $\hat{H}_k$ dans les troisième et quatrième termes de droite de $(1-2)$ par $W_k^0 = diag(\tau_k^0)$ et $H_k^0$, respectivement, où $H_k^0 = X_k^0(X_k^{0'}X_k^0)^{-1}X_k^{0'}$ et $X_k^0 = (W_k^0)^{1/2}X$, et pour simplifier encore nous supposons que dans le vrai modèle les classes sont disjointes telles que les éléments diagonaux de $W_k^0$ sont égaux à 0 ou 1. On a : $n_k^0\hat{\Sigma}_k \sim W_k(\Sigma_k, n_k^0 - m)$ et $E_Y(\hat{\Sigma}_k^{-1}) = d_k^0(\Sigma_k^0)^{-1}$ avec : $d_k^0 = \frac{n_k^0}{n_k^0 - (m + p_k^0 + 1)}$ [3]
On a : $tr[E_Y(\hat{\Sigma}_k^{-1})E_Y\{(\beta_k^* - \hat{\beta}_k)'(X_k^0)'X_k^0(\beta_k^* - \hat{\beta}_k)\}] = d_k^0 p_k^0 m$ (1-3)
et $E_Y\sum_{k=1}^{K^0} tr(W_k^0)tr(\hat{\Sigma}_k^{-1}\Sigma_k^0) = \sum_{k=1}^{K^0} tr(W_k^0)d_k^0 m$ (1-4)
4. En remplaçant (1-3) et (1-4) dans le deuxième et le troisième terme de (1-2) respectivement, et $K^0$ et $W_k^0$ par $K$ et $\hat{W}_k$ respectivement,on obtient un estimateur de $\Delta(\hat{\theta}(Y))$ définie par :
$MRC = \sum_{k=1}^{K}\hat{n}_k log(|\hat{\Sigma}_k|) - 2\sum_{k=1}^{K}\hat{n}_k log(\hat{\alpha}_k) + \sum_{k=1}^{K}\hat{d}_k m(p_k + \hat{n}_k)$
où $\hat{d}_k = \frac{\hat{n}_k}{\hat{n}_k - (m + p_k + 1)}$ et $\hat{n}_k = tr(\hat{W}_k)$

**Remarque :**

- si $m = 1$ (cas univarié), on a :

$$MRC = \sum_{k=1}^{K}\hat{n}_k\log(\hat{\sigma}_k^2) + \sum_{k=1}^{K}\hat{n}_k(\hat{n}_k + p_k)(\hat{n}_k - p_k - 2) - 2\sum_{k=1}^{K}\hat{n}_k\log(\hat{\alpha}_k)$$

qui est le critère défini par Naik et al.(2007)
- Si $K = 1$ et $m = 1$ (cas seule composante et univarié) on a :

$$MRC = nlog(\hat{\sigma}^2) + \frac{n(n+p)}{n-p-2} = AIC_c$$

on retrouve le critère d'Akaike corrigé.

# 3 Dérivation du critère MRC pour un modèle de mélange autorégressif vectoriel

Soit $MVAR(m, K, p_1, p_2, ..., p_k)$ un modèle de mélange autorégressif vectoriel relatif à un vecteur $y_t$ de dimension $m$ avec $K$ composantes dont l'expression est :

$$F(y_t/\mathfrak{F}_{t-1}) = \sum_{k=1}^{K}\alpha_k\Phi[\Sigma_k^{-\frac{1}{2}}(y_t - \Theta_{k1}y_{t-1} - \Theta_{k2}y_{t-2} - ... - \Theta_{kp_k}y_{t-p_k})], t = p+1, ...n \quad (5)$$

où $p = max(p_1, p_2, ..., p_K)$ avec $p_k$ l'ordre du k$^{\text{ième}}$ autorégressif, $\mathfrak{F}_{t-1}$ est la tribu engendrée par $\{y_{t-1}, y_{t-2}, y_{t-3}, ....\}$, $\Phi(.)$ est la distribution normale multivariée de moyenne nulle et de matrice de variances-covariances $I$, $\alpha_k$ est la probabilité de la k$^{\text{ième}}$ composante telle que $\alpha_1 \geq \alpha_2 \geq .... \geq \alpha_k \geq 0$ et $\sum \alpha_i = 1$, $\Theta_{k1}, ..., \Theta_{kp_k}$ sont des matrices d'ordre $m \times m$ , et $\Sigma_k$ est la matrice variance-covariance d'ordre $m \times m$.

## 3.1 Estimation des paramètres

Soit $y_1, ..., y_n$ des observations du modèle $MVAR(m, K, p_1, ..., p_k)$. Pour $t = 1, ..., n$ soit $z_t = (z_{t,1}, ..., z_{t,K})'$ où $z_{t,k} = 1$ si $y_t$ provient de la k$^{\text{ième}}$ composante pour $k = 1, ..., K$ et $z_{t,k} = 0$ sinon, les vecteurs $z_1, ..., z_n$ sont généralement non observés, la densité des données complètes $(y_t, z_t)$ est : $\prod_{k=1}^{K} [\alpha_k f_k(\mathbf{e}_{kt}, \Sigma_k)]^{z_{t,k}}$
où $f_k(\mathbf{e}_{kt}, \Sigma_k) = (2\Pi)^{-\frac{m}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp(-\frac{1}{2}\mathbf{e}'_{kt}\Sigma_k^{-1}\mathbf{e}_{kt})$
, $\mathbf{e}_{kt} = y_t - \Theta_{k1}y_{t-1} - \Theta_{k2}y_{t-2} - ... - \Theta_{kp_k}y_{t-p_k} = y_t - \beta_k x_{kt}$ avec $\beta_k = [\Theta_{k1}, ...., \Theta_{kp_k}]$
et $x_{kt} = (y'_{t-1}, y'_{t-2}, ...., y'_{t-p_k})'$, pour $k = 1, ..., K$. La log-vraissemblance est donnée par :

$$L(\phi, Y, Z) = \sum_{t=p+1}^{n} \{ \sum_{k=1}^{K} z_{t,k} \log(\alpha_k) - \frac{1}{2} \sum_{k=1}^{K} z_{t,k} \log |\Sigma_k| - \frac{1}{2} \sum_{k=1}^{K} z_{t,k} (\mathbf{e}'_{kt} \Sigma_k^{-1} \mathbf{e}_{kt}) \}$$

avec : $\alpha = (\alpha_1, ...., \alpha_{k-1})'$ et $\phi = (\alpha', vec(\beta_1)', vec(\Sigma_1)', ......., vec(\beta_K)', vec(\Sigma_K)')'$
Pour estimer les paramètres on utilise l'algorithme EM, nous avons :

- $\hat{\alpha}_k = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\tau}_{t,k}$
- $\hat{\beta}'_k = (\sum_{t=p+1}^{T} \hat{\tau}_{t,k} x_{kt} x'_{kt})^{-1} (\sum_{t=p+1}^{T} \hat{\tau}_{t,k} x_{kt} y'_{kt})$
- $\hat{\Sigma}_k = \frac{\sum_{t=p+1}^{T} \hat{\tau}_{t,k} \hat{\mathbf{e}}_{kt} \hat{\mathbf{e}}'_{kt}}{\sum_{t=p+1}^{T} \hat{\tau}_{t,k}}$

## 3.2 Dérivation de $MRC$

Soit $Y = (y_{p+1}, ..., y_n)'$, $X = (x_{p+1}, ..., x_n)'$, $x_j = (y'_{j-1}, ..., y'_{j-p_k})'$ et $\beta'_k = [\Theta_{k1}, ...., \Theta_{kp_k}]'$ où $j = p+1, ..., n$ et $k = 1, ..., K$. Alors l'écart des log-vraisemblances du modèle candidat de mélange autogéressif vectoriel et le vrai modèle est :
$\Delta(\hat{\theta}(Y)) = -2E_Y E_{Y^*} L(\hat{\phi}, \hat{\tau}, Y^*, X)$
$\Delta(\hat{\theta}(Y)) = -2E_Y E_{Y^*} [\sum_{k=1}^{K} \sum_{j=p+1}^{n} \hat{\tau}_{j,k} \{\log \hat{\alpha}_k + \log f_k(y_j^*, x_j, \hat{\beta}_k, \hat{\Sigma}_k)\}]$ $\hat{\phi}$, $\hat{\tau}$ et $Y^*$ sont définis dans le paragraphe (1). La dernière expression de $\Delta(\hat{\theta}(Y))$ est la même que celle de (1-1), et par conséquent $\Delta(\hat{\theta}(Y))$ peut être approché par le dernier terme de (1-2). Une application des techniques utilisées dans la démonstration de la proposition de la partie 2.2 entraîne finalement l'expression du $MRC$ pour un modèle de mélange multivarié :

$$MRC = \sum_{k=1}^{K} \hat{n}_k log(|\hat{\Sigma}_k|) - 2 \sum_{k=1}^{K} \hat{n}_k log(\hat{\alpha}_k) + \sum_{k=1}^{K} \hat{d}_k m(p_k + \hat{n}_k)$$

où $\hat{d}_k = \frac{\hat{n}_k}{\hat{n}_k-(m+p_k+1)}$ et $\hat{n}_k = tr(\hat{W}_k)$.

# Conclusion

Le critère $MRC$ est une extension du critère $AIC_c$ dans le cas des modèles de mélange de régression. Nous avons proposé une démonstration du critère $MRC$ dans le cas des modèles de mélange de régression multivariée. Nous avons aussi donnée une formulation du critère $MRC$ pour sélectionner les modèles de mélange autorégressifs vectoriels.

# Bibliographie

[1] Akaike, H.(1985),Prediction and entropy,*In A Celebration of Statistics*,Ed.A.C.Atkinson and S.E.Fienberg,pp.1–24,New York :Springer

[2] Dempster,A.P, Laird,N.M and Rubin,D.P. (1977) Maximum likelihood from incomplete data via the EM algorithm.*Journal of the Royal Statistical Society Series B*, 39,1–38.

[3] Edward J.Bedrick and Chih-Ling Tsai. (1994) Model Selection for Multivariate Regression in small Samples ,*Biometrics*, 50, 226–231.

[4] Leroux,B.(1992),Consistent estimation of a mixing distribution,*The Annals of Statistics*,20,1350–1360

[5] Naik PA,Shi P,Tsai CL. (2007) Extending the Akaike information criterion to mixture regression models.*J Am Stat Assoc* , 102(477) :244–254.

[6] P.W.Fong,W.K.Li,C.W.Yau and C.S.Wong (2007) On a Mixture Vector Autoregressive Model. *The canadian Journal of Statisyics / la Revue canadienne de Statistique*,14 1080–10100

[7] Schawarz G .(1978) Srivastava,M.S.,and Khatri,C.G.(1979) An Introduction to Multivariate Statistics.North-Holland,New York.