
Caractérisation de mélange de distributions gaussiennes

Didier MAQUIN, Gilles MOUROT et José RAGOT

Institut National Polytechnique de Lorraine

Centre de Recherche en Automatique de Nancy

UMR 7039 CNRS – Université Henri Poincaré, Nancy 1 – INPL

Contexte de l'étude

Contexte de l'étude

- Description d'un comportement ou d'une propriété à l'aide de plusieurs modèles et d'un mécanisme de commutation ou de transition
- Notion de commutation de modèles, de multi-modèle, de mélange de modèles, etc.

Contexte de l'étude

- Description d'un comportement ou d'une propriété à l'aide de plusieurs modèles et d'un mécanisme de commutation ou de transition
- Notion de commutation de modèles, de multi-modèle, de mélange de modèles, etc.

Difficultés – Avantages

Contexte de l'étude

- Description d'un comportement ou d'une propriété à l'aide de plusieurs modèles et d'un mécanisme de commutation ou de transition
- Notion de commutation de modèles, de multi-modèle, de mélange de modèles, etc.

Difficultés – Avantages

- Compromis entre le nombre de modèles "locaux" et leur complexité
- Transposition de résultats classiques dans le cas de modèles locaux élémentaires
- Problème de classification des données

Distributions de variables aléatoires

Distribution unique

Distribution unique

- Densité de probabilité “locale” : $p_j(y|\theta_j)$
- Paramètres de la distribution “locale” : θ_j

Distributions de variables aléatoires

Distribution unique

- Densité de probabilité “locale” : $p_j(y|\theta_j)$
- Paramètres de la distribution “locale” : θ_j

Distribution mélange

Distribution unique

- Densité de probabilité “locale” : $p_j(y|\theta_j)$
- Paramètres de la distribution “locale” : θ_j

Distribution mélange

- Densité de probabilité “globale” :

$$p(y|\alpha, \theta) = \sum_{j=1}^M \alpha_j p_j(y|\theta_j)$$

- Paramètres de mélange : $0 < \alpha_j < 1$, $\sum_{j=1}^M \alpha_j = 1$

Distributions de variables aléatoires

Distribution unique

- Densité de probabilité “locale” : $p_j(y|\theta_j)$
- Paramètres de la distribution “locale” : θ_j

Distribution mélange

- Densité de probabilité “globale” :

$$p(y|\alpha, \theta) = \sum_{j=1}^M \alpha_j p_j(y|\theta_j)$$

- Paramètres de mélange : $0 < \alpha_j < 1$, $\sum_{j=1}^M \alpha_j = 1$

Questions

Distribution unique

- Densité de probabilité “locale” : $p_j(y|\theta_j)$
- Paramètres de la distribution “locale” : θ_j

Distribution mélange

- Densité de probabilité “globale” :

$$p(y|\alpha, \theta) = \sum_{j=1}^M \alpha_j p_j(y|\theta_j)$$

- Paramètres de mélange : $0 < \alpha_j < 1$, $\sum_{j=1}^M \alpha_j = 1$

Questions

- Structure des lois “locales”
- Nombre de modèles locaux

Lois normales

$$p_1(y|\theta_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{y - m_1}{\sigma_1}\right)^2\right) \quad \theta_1 = (m_1 \quad \sigma_1)^T$$

$$p_2(y|\theta_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{y - m_2}{\sigma_2}\right)^2\right) \quad \theta_2 = (m_2 \quad \sigma_2)^T$$

Lois normales

$$p_1(y|\theta_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{y - m_1}{\sigma_1}\right)^2\right) \quad \theta_1 = (m_1 \quad \sigma_1)^T$$

$$p_2(y|\theta_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\left(\frac{y - m_2}{\sigma_2}\right)^2\right) \quad \theta_2 = (m_2 \quad \sigma_2)^T$$

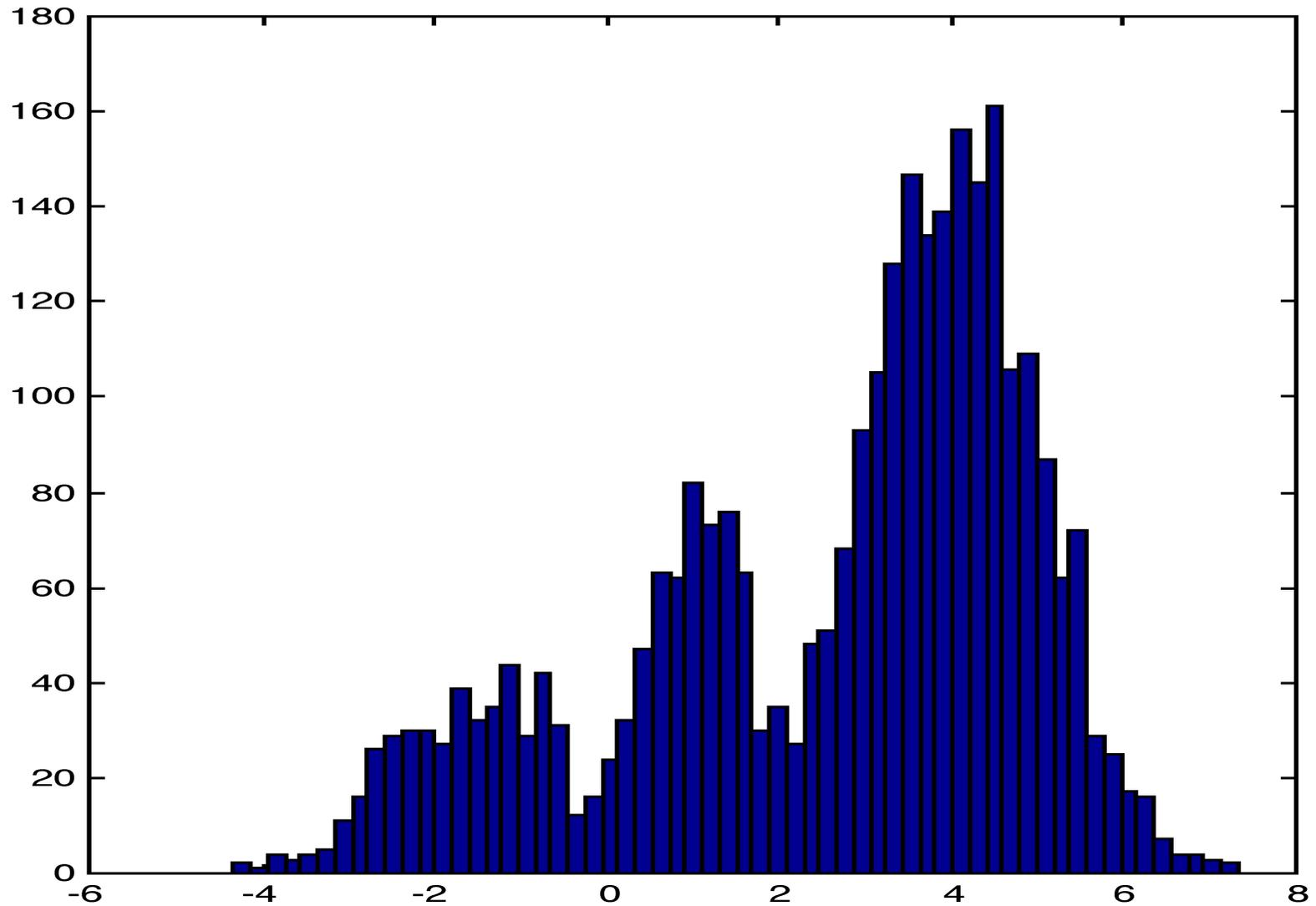
Mélange

$$p(y|\alpha, \theta) = \alpha p_1(y|\theta_1) + (1 - \alpha)p_2(y|\theta_2)$$

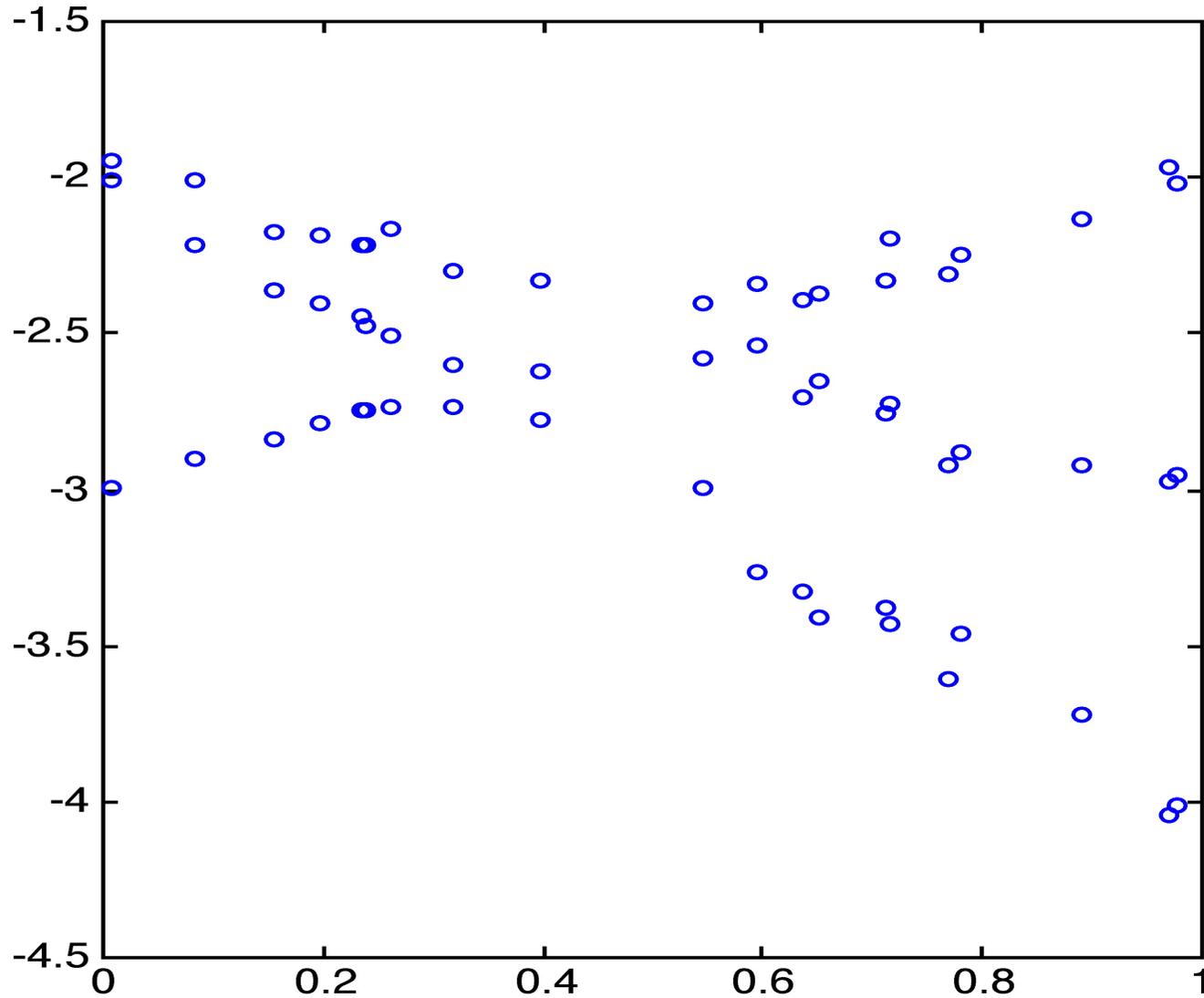
avec

$$\theta = (m_1 \quad m_2 \quad \sigma_1 \quad \sigma_2)^T$$

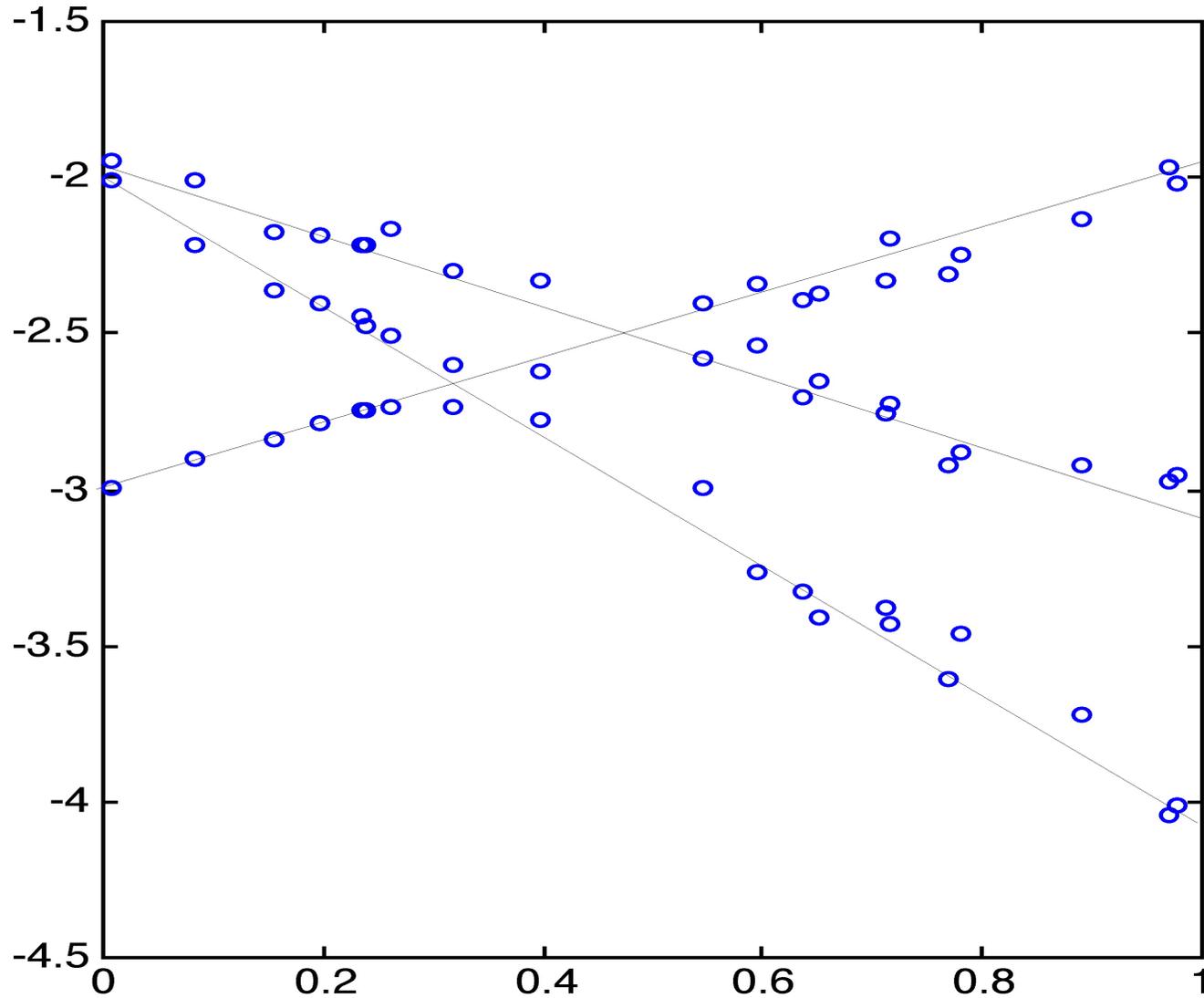
Mélange de distributions normales



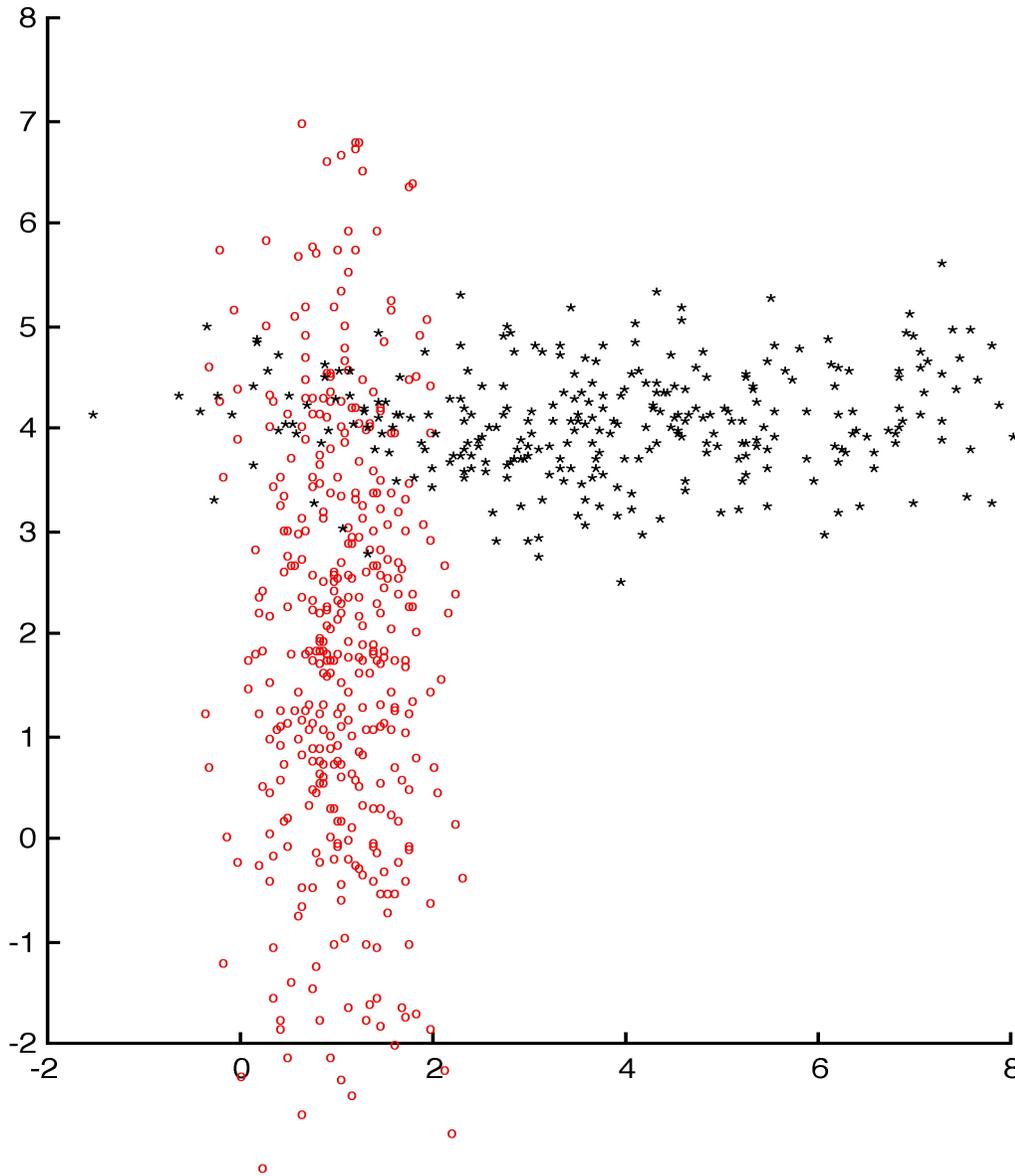
Mélange de modèles affines



Mélange de modèles affines



Lois normales bidimensionnelles



Mélange de M modèles

$$p(y|\alpha, \theta) = \sum_{j=1}^M \alpha_j p(y|\theta_j)$$

Formulation du problème (1)

Mélange de M modèles

$$p(y|\alpha, \theta) = \sum_{j=1}^M \alpha_j p(y|\theta_j)$$

Séquence Y de N observations y_i

$$p(Y|\theta) = \prod_{i=1}^N p(y_i|\alpha, \theta)$$

Formulation du problème (1)

Mélange de M modèles

$$p(y|\alpha, \theta) = \sum_{j=1}^M \alpha_j p(y|\theta_j)$$

Séquence Y de N observations y_i

$$p(Y|\theta) = \prod_{i=1}^N p(y_i|\alpha, \theta)$$

Fonction de log-vraisemblance des observations

$$L(\Psi|Y) = \sum_{i=1}^N \ln \left(\sum_{j=1}^M \alpha_j p(y_i|\theta_j) \right)$$

Formulation du problème (2)

Fonction de log-vraisemblance des observations

$$L(\Psi|Y) = \sum_{i=1}^N \ln \left(\sum_{j=1}^M \alpha_j p(y_i|\theta_j) \right)$$

$$\Psi = (\alpha_1 \dots \alpha_M \quad \theta_1 \dots \theta_M)^T$$

Formulation du problème (2)

Fonction de log-vraisemblance des observations

$$L(\Psi|Y) = \sum_{i=1}^N \ln \left(\sum_{j=1}^M \alpha_j p(y_i|\theta_j) \right)$$

$$\Psi = (\alpha_1 \dots \alpha_M \quad \theta_1 \dots \theta_M)^T$$

Estimation au sens du maximum de vraisemblance

$$\hat{\Psi} = \arg \max_{\psi} \sum_{i=1}^N \ln \left(\sum_{j=1}^M \alpha_j p(y_i|\theta_j) \right)$$

sous la contrainte $\sum_{j=1}^M \alpha_j = 1$

Lagrangien

$$\Phi = \sum_{i=1}^N \ln \left(\sum_{j=1}^M \alpha_j p(y_i | \theta_j) \right) + \lambda \left(\sum_{j=1}^M \alpha_j - 1 \right)$$

Lagrangien

$$\Phi = \sum_{i=1}^N \ln \left(\sum_{j=1}^M \alpha_j p(y_i | \theta_j) \right) + \lambda \left(\sum_{j=1}^M \alpha_j - 1 \right)$$

Equations d'optimalité (1er ordre)

$$\frac{\partial \Phi}{\partial \alpha_j} \Big|_{\Psi = \hat{\Psi}} = 0 \quad \frac{\partial \Phi}{\partial \theta_j} \Big|_{\Psi = \hat{\Psi}} = 0 \quad \frac{\partial \Phi}{\partial \lambda} \Big|_{\Psi = \hat{\Psi}} = 0$$

Lagrangien

$$\Phi = \sum_{i=1}^N \ln \left(\sum_{j=1}^M \alpha_j p(y_i | \theta_j) \right) + \lambda \left(\sum_{j=1}^M \alpha_j - 1 \right)$$

Equations d'optimalité (1er ordre)

$$\left. \frac{\partial \Phi}{\partial \alpha_j} \right|_{\Psi = \hat{\Psi}} = 0 \quad \left. \frac{\partial \Phi}{\partial \theta_j} \right|_{\Psi = \hat{\Psi}} = 0 \quad \left. \frac{\partial \Phi}{\partial \lambda} \right|_{\Psi = \hat{\Psi}} = 0$$

Notations

$$p(y_i | \theta_j) = p_{i,j} \quad \sum_{j=1}^M \alpha_j p(y_i | \theta_j) = p_i$$

Système à résoudre

$$\sum_{i=1}^N \frac{p_{i,j}}{\sum_{j=1}^M \alpha_j p_{i,j}} + \lambda = 0$$

$$\sum_{i=1}^N \frac{\alpha_j \frac{\partial p_{i,j}}{\partial \theta_j}}{\sum_{j=1}^M \alpha_j p_{i,j}} = 0$$

$$\sum_{j=1}^M \alpha_j - 1 = 0$$

Systeme à résoudre

$$\sum_{i=1}^N \frac{p_{i,j}}{\sum_{j=1}^M \alpha_j p_{i,j}} + \lambda = 0$$

$$\sum_{i=1}^N \frac{\alpha_j \frac{\partial p_{i,j}}{\partial \theta_j}}{\sum_{j=1}^M \alpha_j p_{i,j}} = 0$$

$$\sum_{j=1}^M \alpha_j - 1 = 0$$

Résolution hiérarchisée

Equations implicites (cas de deux modèles)

Equations implicites (cas de deux modèles)

$$m_j = \frac{\sum_{i=1}^N \frac{p_{j,i}}{p_i} y_i}{\sum_{i=1}^N \frac{p_{j,i}}{p_i}}$$
$$\sigma_j^2 = \frac{\sum_{i=1}^N \frac{p_{j,i}}{p_i} (y_i - m_j)^2}{\sum_{i=1}^N \frac{p_{j,i}}{p_i}}$$

Equations implicites (cas de deux modèles)

$$m_j = \frac{\sum_{i=1}^N \frac{p_{j,i}}{p_i} y_i}{\sum_{i=1}^N \frac{p_{j,i}}{p_i}} \quad \sigma_j^2 = \frac{\sum_{i=1}^N \frac{p_{j,i}}{p_i} (y_i - m_j)^2}{\sum_{i=1}^N \frac{p_{j,i}}{p_i}}$$

$$p_{j,i} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2} \left(\frac{y_i - m_j}{\sigma_j}\right)^2\right)$$

Equations implicites (cas de deux modèles)

$$m_j = \frac{\sum_{i=1}^N \frac{p_{j,i}}{p_i} y_i}{\sum_{i=1}^N \frac{p_{j,i}}{p_i}} \quad \sigma_j^2 = \frac{\sum_{i=1}^N \frac{p_{j,i}}{p_i} (y_i - m_j)^2}{\sum_{i=1}^N \frac{p_{j,i}}{p_i}}$$

$$p_{j,i} = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{1}{2} \left(\frac{y_i - m_j}{\sigma_j}\right)^2\right)$$

$$\sum_{i=1}^N \frac{1}{\alpha + \frac{p_{2,i}}{p_{1,i} - p_{2,i}}} = 0$$

Algorithme à itération directe

Algorithme à itération directe

$$m_j^{(k+1)} = \frac{1}{N} \sum_{i=1}^N \frac{p_{j,i}^{(k)}}{p_i^{(k)}} y_i$$

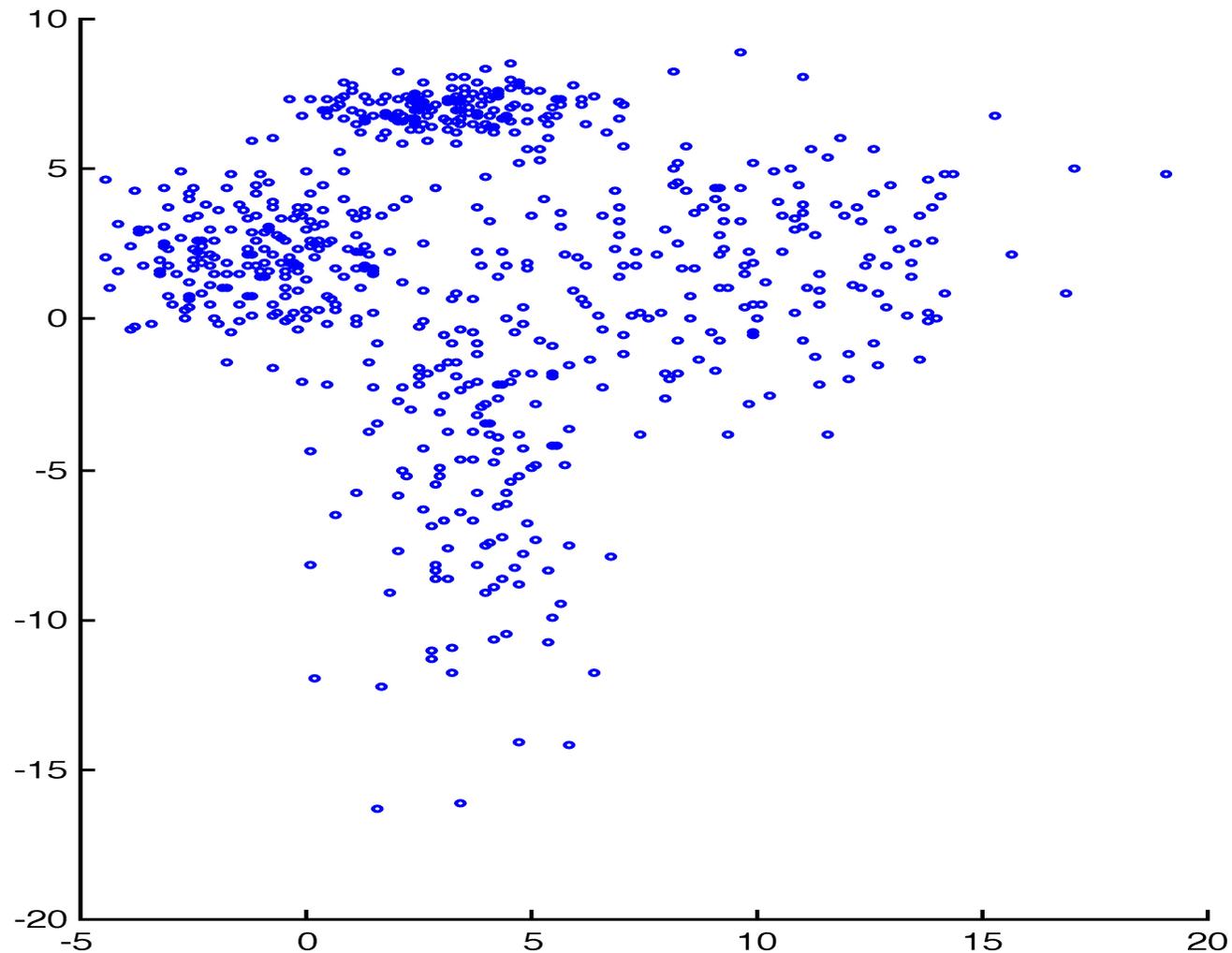
$$(\sigma_j^2)^{(k+1)} = \frac{1}{N} \sum_{i=1}^N \frac{p_{j,i}^{(k)}}{p_i^{(k)}} \left(y_i - m_j^{(k)} \right)^2$$

Algorithme à itération directe

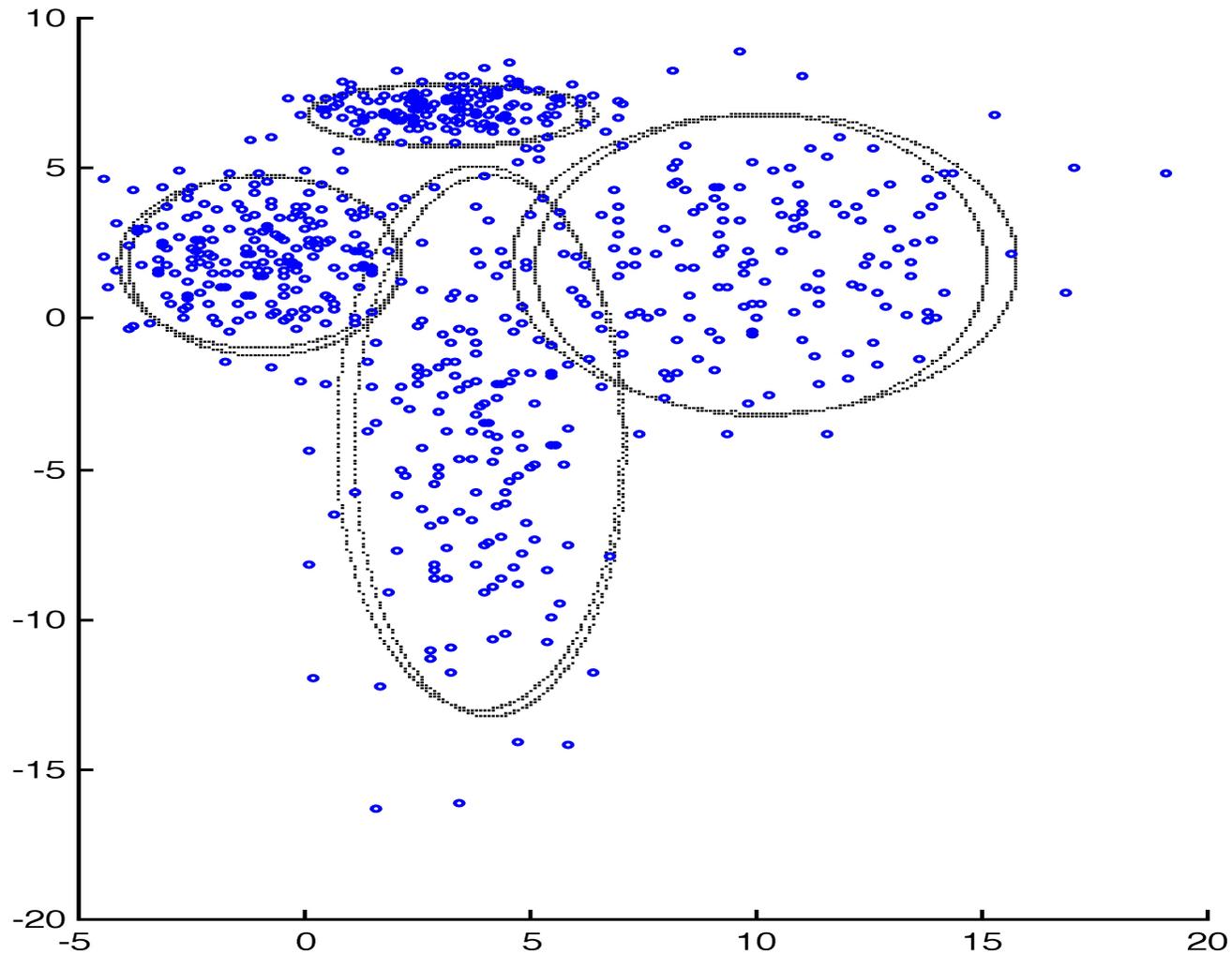
$$m_j^{(k+1)} = \frac{1}{N} \sum_{i=1}^N \frac{p_{j,i}^{(k)}}{p_i^{(k)}} y_i \quad (\sigma_j^2)^{(k+1)} = \frac{1}{N} \sum_{i=1}^N \frac{p_{j,i}^{(k)}}{p_i^{(k)}} \left(y_i - m_j^{(k)} \right)^2$$

$$\alpha^{(k+1)} = \alpha^{(k)} + \frac{\sum_{i=1}^N \frac{1}{\alpha^{(k)} + \frac{p_{2,i}^{(k)}}{p_{1,i}^{(k)} - p_{2,i}^{(k)}}}}{\sum_{i=1}^N \frac{1}{\left(\alpha^{(k)} + \frac{p_{2,i}^{(k)}}{p_{1,i}^{(k)} - p_{2,i}^{(k)}} \right)^2}}$$

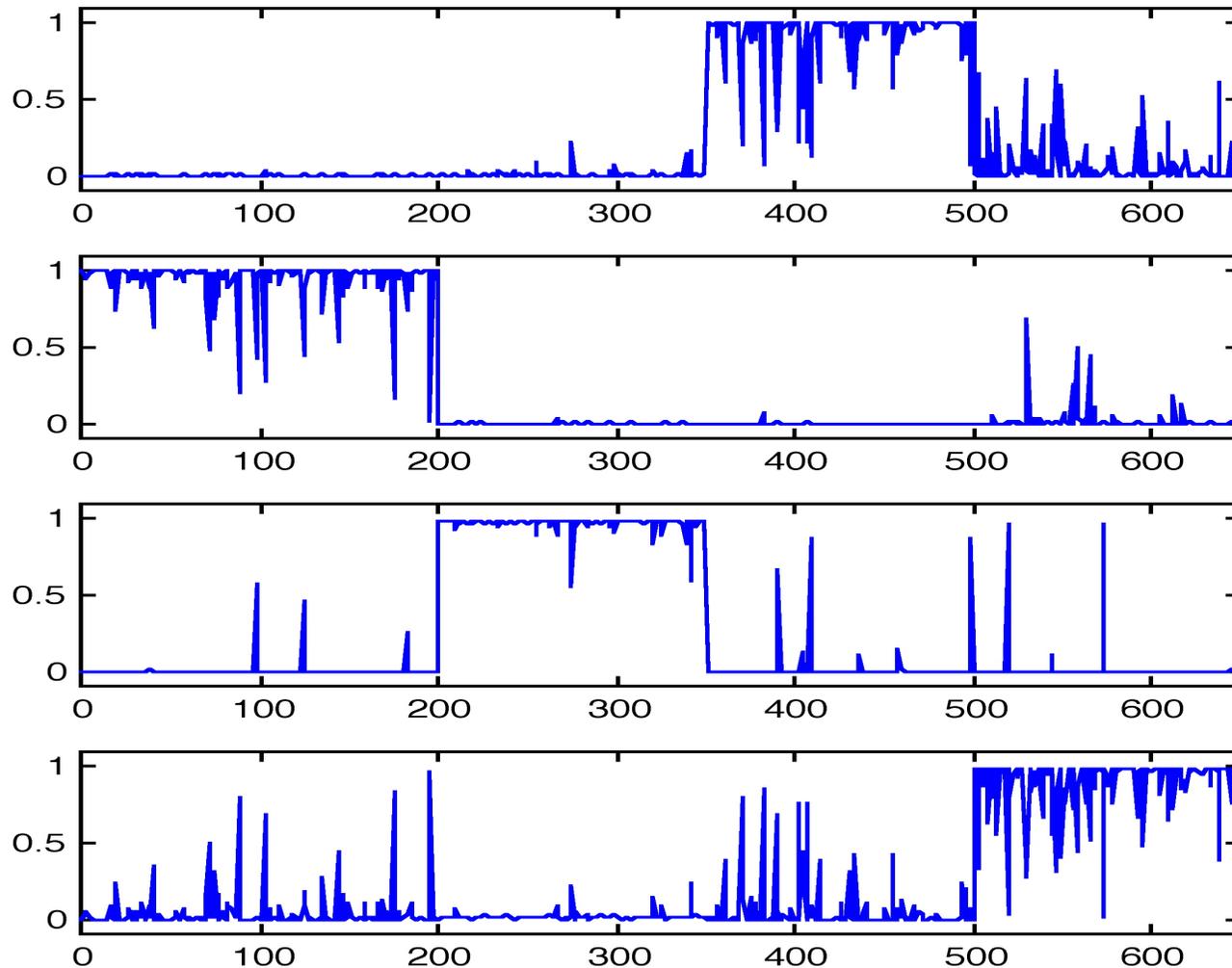
Mélange de 4 distributions normales



Ellipsoïdes théoriques et estimés (1 écart-type)



Fonction d'appartenance des observations aux modèles



Caractéristiques “vraies”

Distribution	Centre		Ecart-type		Effectif
1	-1.00	2.00	1.50	1.50	200
2	3.00	7.00	1.50	0.50	150
3	9.00	2.00	2.50	2.50	150
4	4.00	-4.00	1.50	4.50	150

Caractéristiques estimées

Distribution	Centre		Ecart-type		Proportion
1	-1.08	2.03	1.53	1.50	29.8
2	3.16	6.89	1.43	0.58	23.6
3	9.19	1.34	2.53	2.37	23.6
4	3.87	-3.99	1.55	4.06	23.0

- Intérêt de la représentation multimodèle des systèmes NL

- Intérêt de la représentation multimodèle des systèmes NL
- Capacité à transposer des résultats “classiques” (cas linéaire) aux systèmes NL

- Intérêt de la représentation multimodèle des systèmes NL
- Capacité à transposer des résultats “classiques” (cas linéaire) aux systèmes NL
- Portée générale de la méthode proposée liée au pouvoir de représentation des multimodèles

- Intérêt de la représentation multimodèle des systèmes NL
- Capacité à transposer des résultats “classiques” (cas linéaire) aux systèmes NL
- Portée générale de la méthode proposée liée au pouvoir de représentation des multimodèles
- Application prometteuse dans le contexte du cryptage

- Optimisation du nombre de modèles locaux

- Optimisation du nombre de modèles locaux
- Compromis précision – complexité

- Optimisation du nombre de modèles locaux
- Compromis précision – complexité
- Prise en compte des valeurs aberrantes (aspect robustesse)

- Optimisation du nombre de modèles locaux
- Compromis précision – complexité
- Prise en compte des valeurs aberrantes (aspect robustesse)
- Allocation booléenne des données aux modes de fonctionnement