

# Modèles bayésiens pour l'apprentissage non supervisé

François Caron

INRIA Bordeaux Sud Ouest  
Equipe CQFD

Jeudi 23 octobre 2008

# Sommaire

- 1 Introduction
- 2 Modèles bayésiens non paramétriques
- 3 Algorithme MCMC
- 4 Classification non supervisée dynamique
- 5 Conclusion

# Sommaire

- 1 Introduction
- 2 Modèles bayésiens non paramétriques
- 3 Algorithme MCMC
- 4 Classification non supervisée dynamique
- 5 Conclusion

# Apprentissage non supervisé

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$

# Apprentissage non supervisé

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Estimer un paramètre inconnu  $\theta$  à partir des données

# Apprentissage non supervisé

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Estimer un paramètre inconnu  $\theta$  à partir des données
- Réduction de dimension, classification . . .

# Apprentissage non supervisé

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Estimer un paramètre inconnu  $\theta$  à partir des données
- Réduction de dimension, classification . . .
- Objectif : prédiction, interprétabilité, compression. . .

# Apprentissage non supervisé

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Estimer un paramètre inconnu  $\theta$  à partir des données
- Réduction de dimension, classification . . .
- Objectif : prédiction, interprétabilité, compression. . .
- Approche bayésienne

# Apprentissage non supervisé

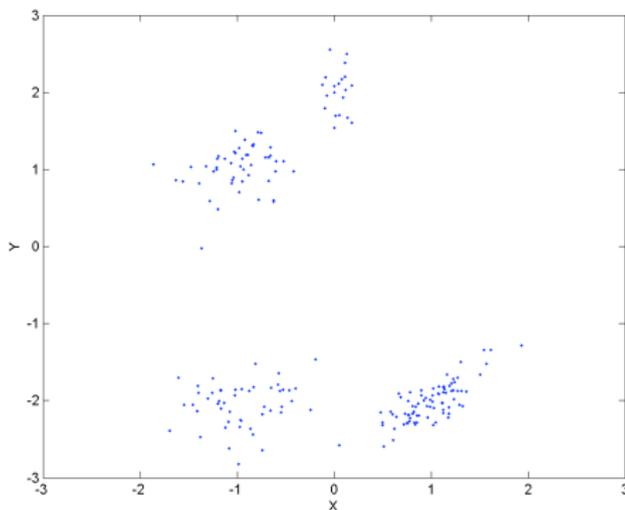
- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Estimer un paramètre inconnu  $\theta$  à partir des données
- Réduction de dimension, classification . . .
- Objectif : prédiction, interprétabilité, compression. . .
- Approche bayésienne
  - ▶ A priori sur le paramètre inconnu  $\theta$

# Apprentissage non supervisé

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Estimer un paramètre inconnu  $\theta$  à partir des données
- Réduction de dimension, classification . . .
- Objectif : prédiction, interprétabilité, compression . . .
- Approche bayésienne
  - ▶ A priori sur le paramètre inconnu  $\theta$
  - ▶ Distribution a posteriori de  $\theta$  conditionnellement aux données

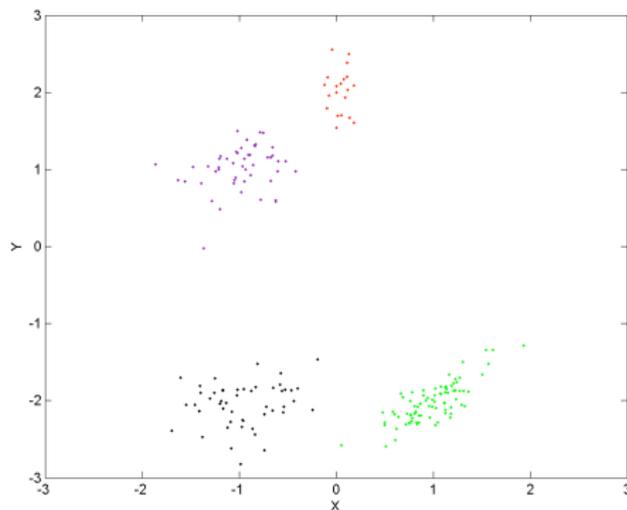
# Classification non supervisée

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Classer ces données en groupes partageant des caractéristiques similaires
- Estimer le centroid de chaque groupe



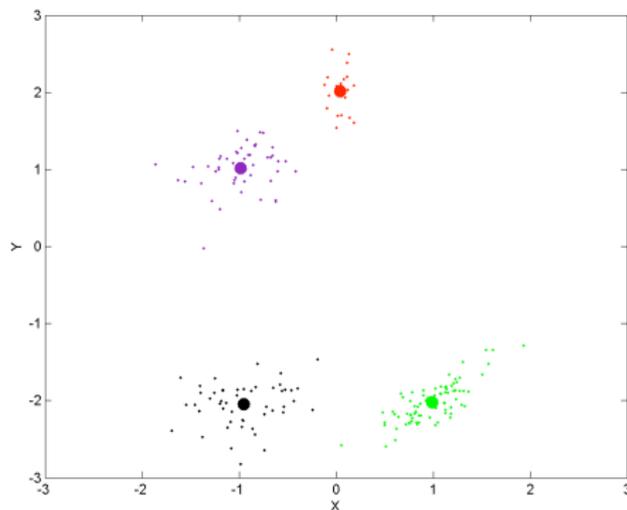
# Classification non supervisée

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Classer ces données en groupes partageant des caractéristiques similaires
- Estimer le centroid de chaque groupe



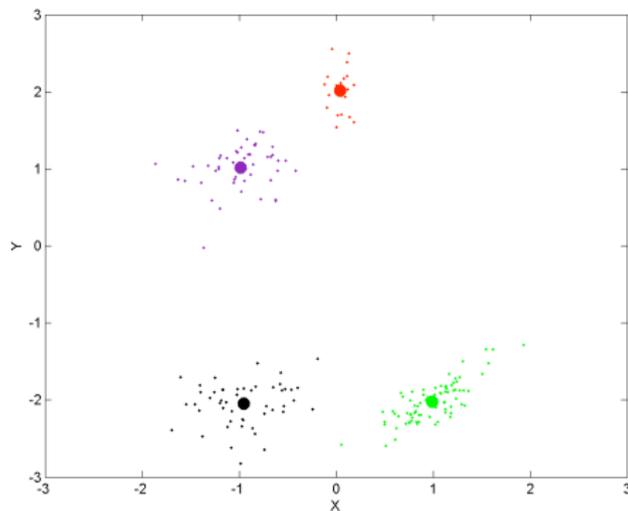
# Classification non supervisée

- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Classer ces données en groupes partageant des caractéristiques similaires
- Estimer le centroid de chaque groupe



# Classification non supervisée

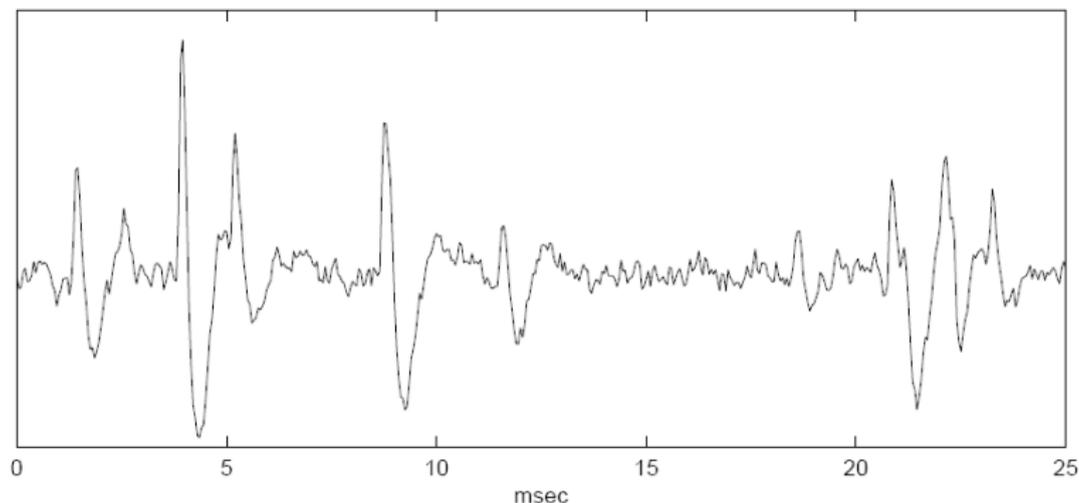
- Ensemble de données  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$
- Pour chaque donnée  $k$ ,  $k = 1, \dots, n$ , estimer la variable de classe  $c_k = j$ ,  $j = 1, 2, \dots$
- Estimer le centroid  $U_j$ ,  $j = 1, 2, \dots$  de chaque groupe



# Exemple : Classification de potentiels d'action

Lewicki, Network : Comput. Neural Syst., 1998

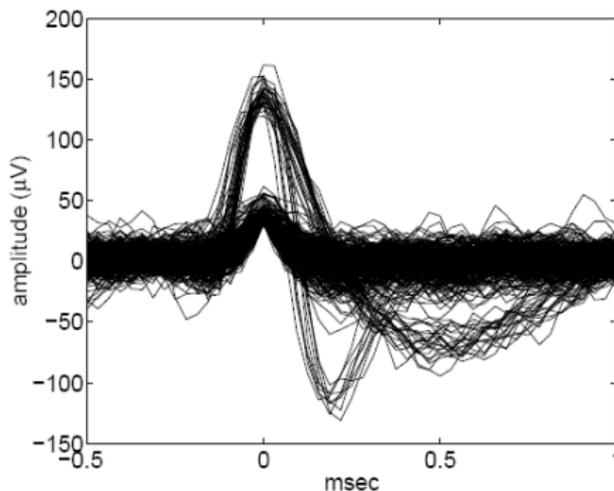
- Brefs potentiels d'action enregistrés par une micro-électrode
- Objectif : classer les signaux afin d'attribuer chaque potentiel à un neurone particulier
- Nombre de neurones inconnu, bruit de fond



# Exemple : Classification de potentiels d'action

Lewicki, Network : Comput. Neural Syst., 1998

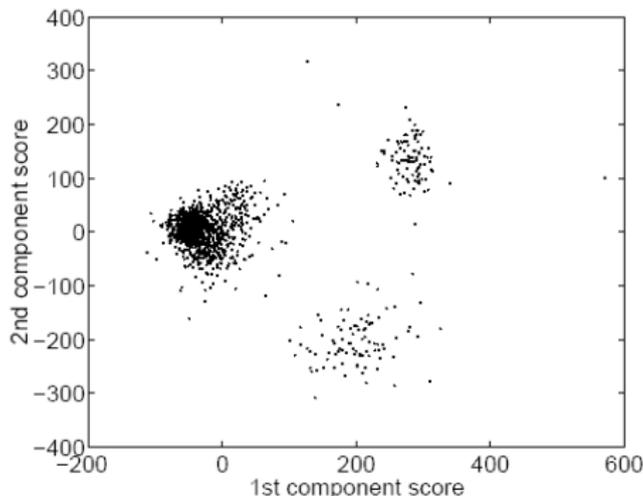
- Brefs potentiels d'action enregistrés par une micro-électrode
- Objectif : classer les signaux afin d'attribuer chaque potentiel à un neurone particulier
- Nombre de neurones inconnu, bruit de fond



# Exemple : Classification de potentiels d'action

Lewicki, Network : Comput. Neural Syst., 1998

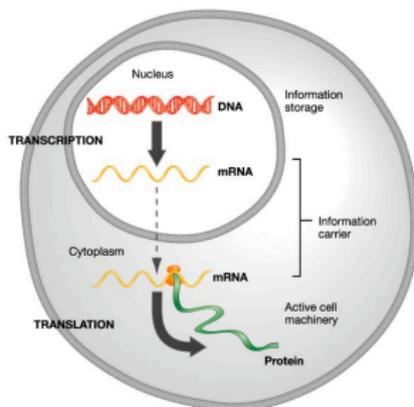
- Brefs potentiels d'action enregistrés par une micro-électrode
- Objectif : classer les signaux afin d'attribuer chaque potentiel à un neurone particulier
- Nombre de neurones inconnu, bruit de fond



# Exemple : Classification de séquences d'expression de gènes

Medvedovic and Sivaganesan, Bioinformatics, 2002.

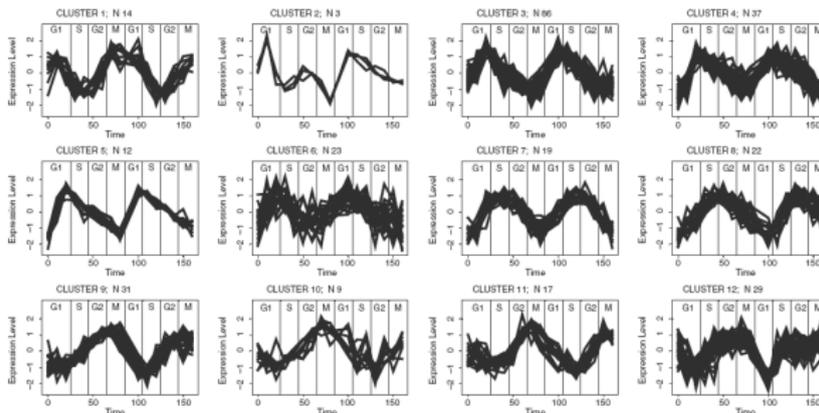
- Puces à ADN : données d'expression de gène pour plusieurs gènes et plusieurs conditions expérimentales
- Classification de gènes avec des séquences d'expression similaires



# Exemple : Classification de séquences d'expression de gènes

Medvedovic and Sivaganesan, Bioinformatics, 2002.

- Puces à ADN : données d'expression de gène pour plusieurs gènes et plusieurs conditions expérimentales
- Classification de gènes avec des séquences d'expression similaires



# Classification non supervisée

- Nombre de partitions pour  $n$  données : nombre de Bell

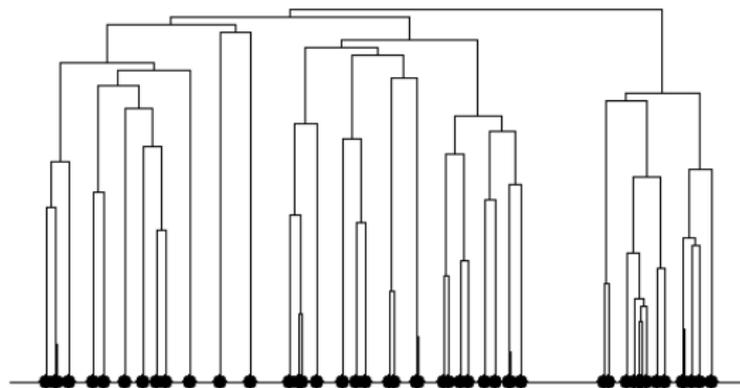
Nombre de données	2	5	10	100	200
Nombre de Bell	2	52	115975	$4.7 \times 10^{115}$	$6 \times 10^{275}$

- Nombre de données important : mise à l'échelle
- Bruit de fond
- Dimension des données
- Autres paramètres/hyperparamètres inconnus
- Nombre de groupes inconnu

# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

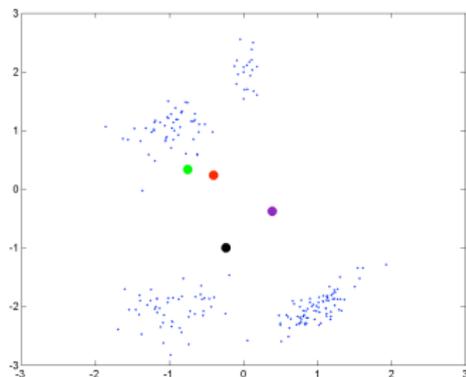
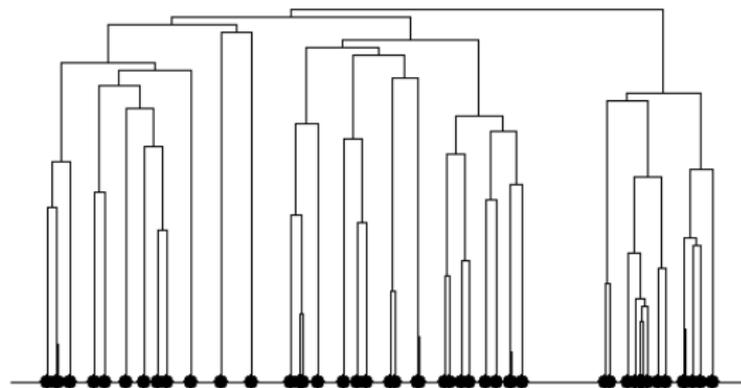
- Clustering hiérarchique



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

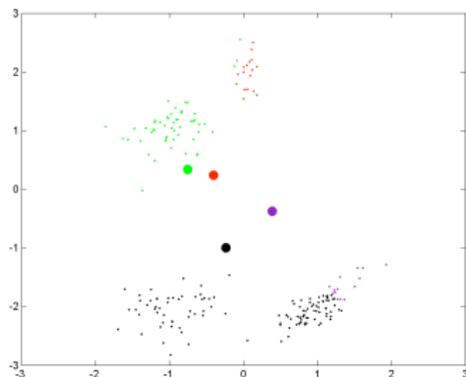
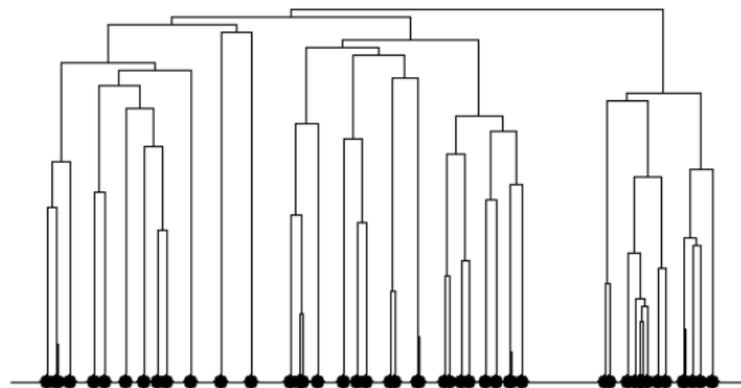
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

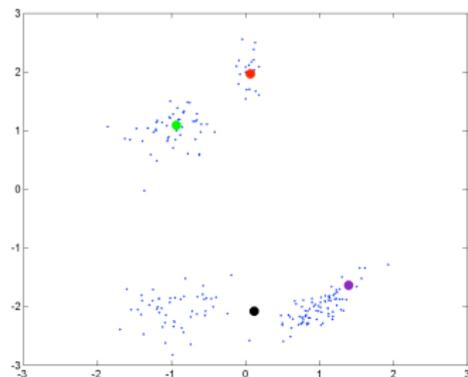
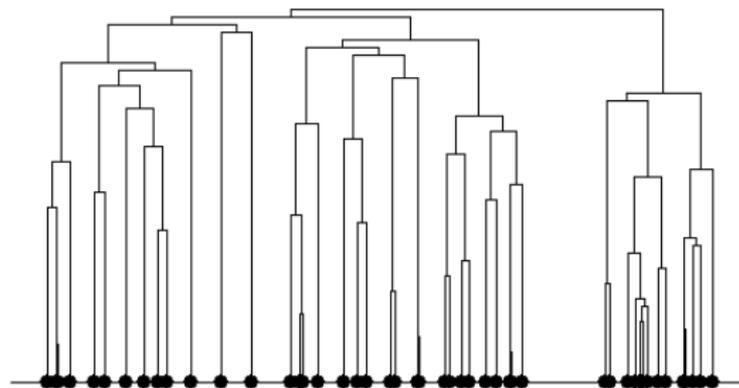
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

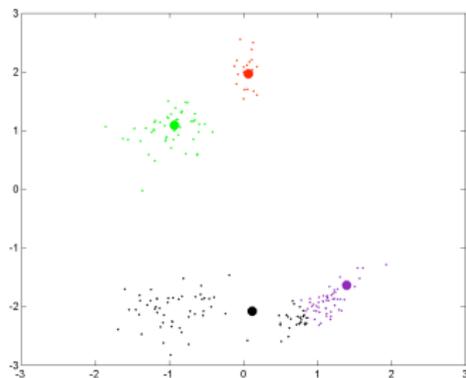
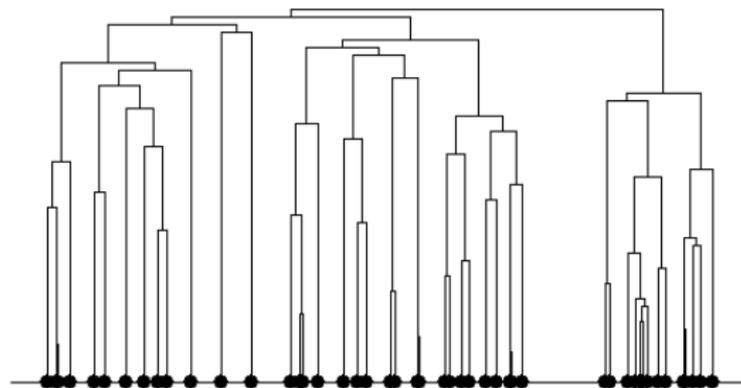
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

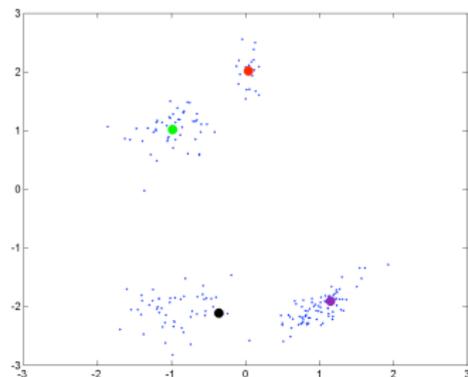
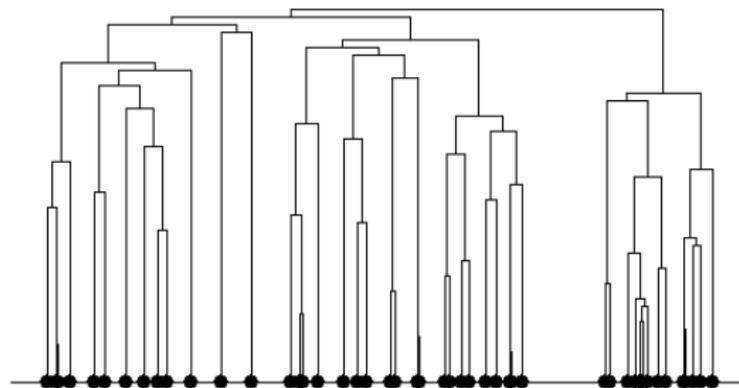
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

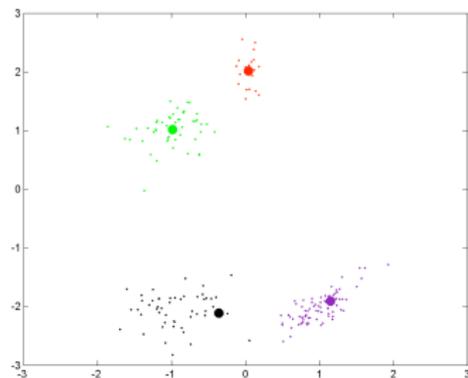
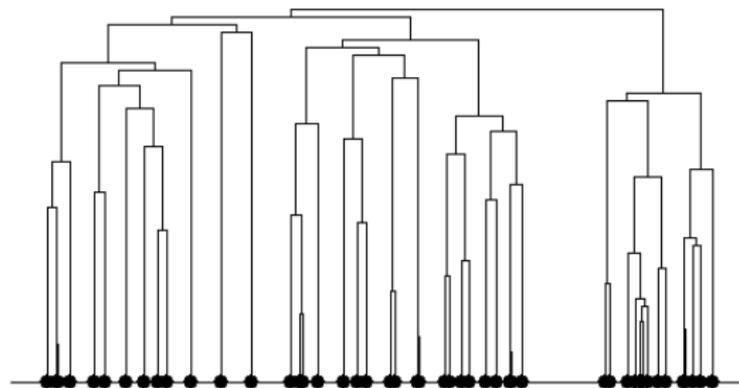
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

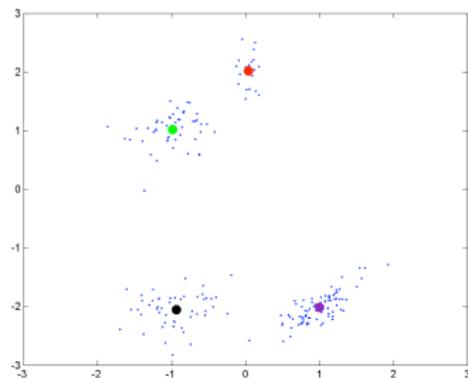
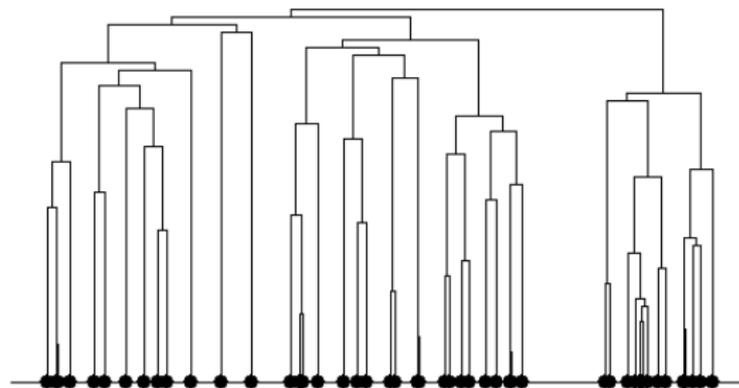
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

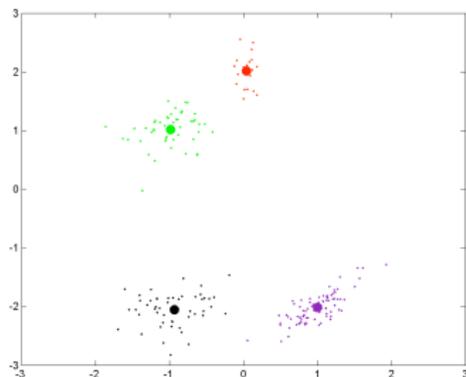
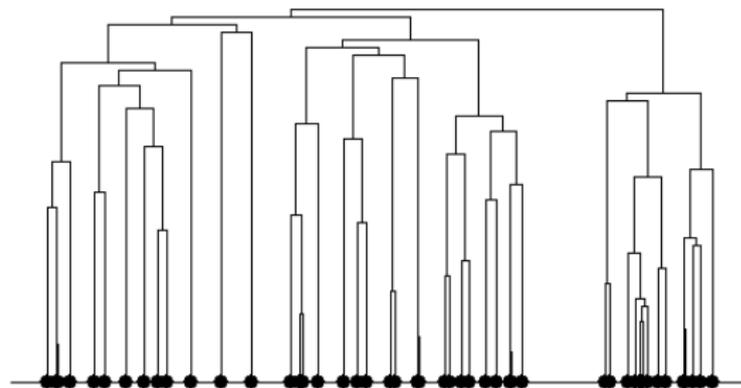
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

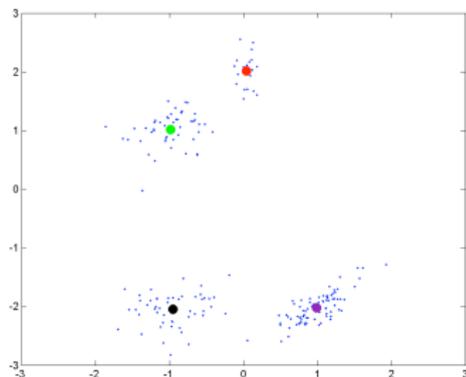
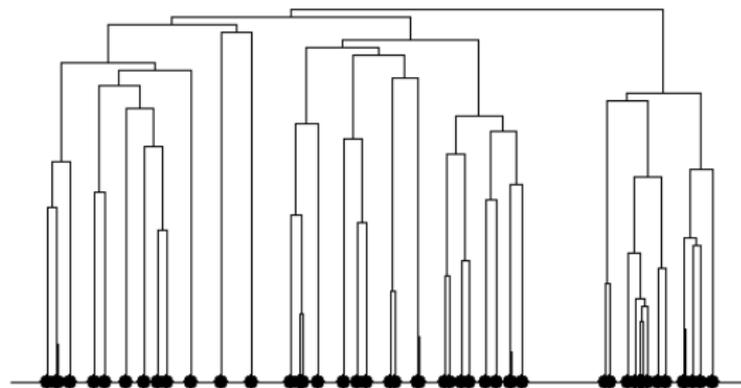
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

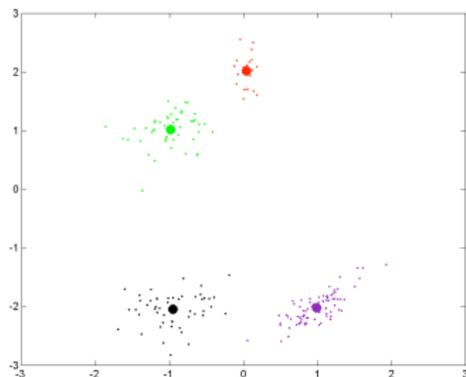
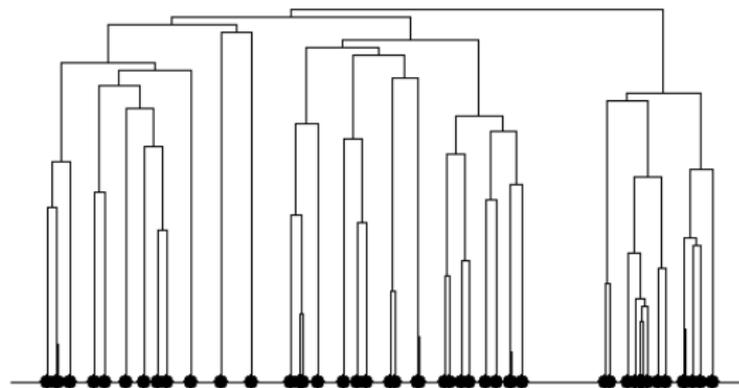
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

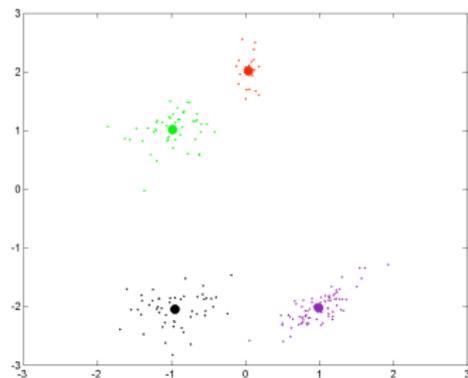
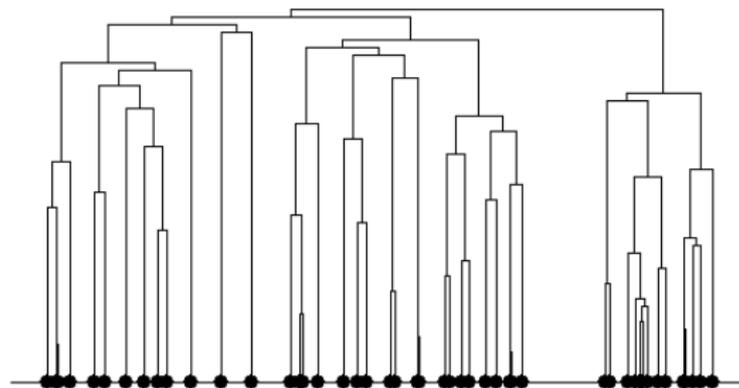
- Clustering hiérarchique
- K-means



# Algorithmes de clustering

Johnson, 1967. MacQueen, 1967.

- Clustering hiérarchique
- K-means
- Clustering basé sur un modèle
  - ▶ Approche bayésienne (empirique)
  - ▶ Paramétrique/non paramétrique

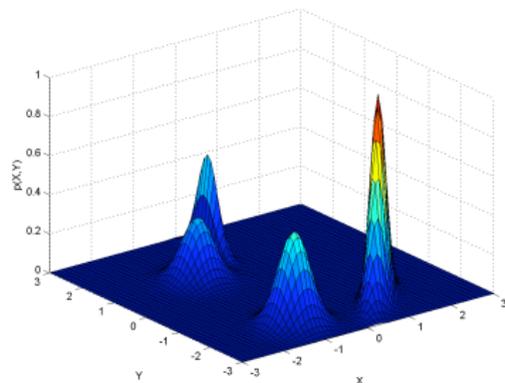
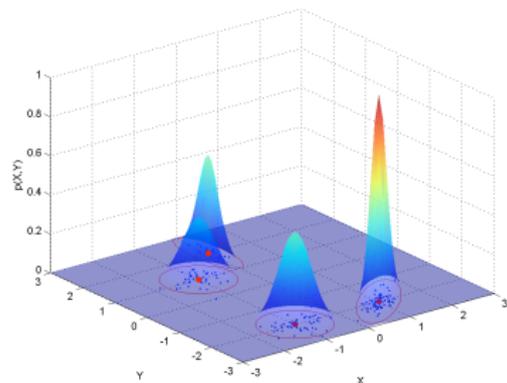


# Clustering basé sur un modèle

- Reformuler comme un problème d'estimation densité
- Données  $\mathbf{z}_1, \dots, \mathbf{z}_n$  i.i.d. selon une distribution inconnue  $F$
- Exemple : Mélange de distributions gaussiennes

$$F = \sum_{j=1}^K \pi_j \mathcal{N}(\cdot | \mu_j, \Sigma_j)$$

où  $\sum_j \pi_j = 1$



# Classification basée sur un modèle

Dempster et al., 1977

- $F$  est supposé être un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$

# Classification basée sur un modèle

Dempster et al., 1977

- $F$  est supposé être un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$
- $U_1, \dots, U_K$  sont les centroïdes des clusters

# Classification basée sur un modèle

Dempster et al., 1977

- $F$  est supposé être un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$
- $U_1, \dots, U_K$  sont les centroïdes des clusters
- $\pi_1, \dots, \pi_K$  sont les tailles relatives des clusters ( $\sum_{j=1}^K \pi_j = 1$ ) de telle façon que  $\Pr(c_k = j) = \pi_j$

# Classification basée sur un modèle

Dempster et al., 1977

- $F$  est supposé être un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$
- $U_1, \dots, U_K$  sont les centroïdes des clusters
- $\pi_1, \dots, \pi_K$  sont les tailles relatives des clusters ( $\sum_{j=1}^K \pi_j = 1$ ) de telle façon que  $\Pr(c_k = j) = \pi_j$
- $f(\cdot)$  est la forme paramétrique d'un cluster (e.g. gaussienne de moyenne et matrice de covariance inconnues)

# Classification basée sur un modèle

Dempster et al., 1977

- $F$  est supposé être un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$
- $U_1, \dots, U_K$  sont les centroïdes des clusters
- $\pi_1, \dots, \pi_K$  sont les tailles relatives des clusters ( $\sum_{j=1}^K \pi_j = 1$ ) de telle façon que  $\Pr(c_k = j) = \pi_j$
- $f(\cdot)$  est la forme paramétrique d'un cluster (e.g. gaussienne de moyenne et matrice de covariance inconnues)
- Objectif : Estimer  $\theta = \{U_{1:K}, \pi_{1:K}\}$  étant donné  $\mathbf{z}_{1:n}$

# Classification basée sur un modèle

Dempster et al., 1977

- $F$  est supposé être un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$
- $U_1, \dots, U_K$  sont les centroïdes des clusters
- $\pi_1, \dots, \pi_K$  sont les tailles relatives des clusters ( $\sum_{j=1}^K \pi_j = 1$ ) de telle façon que  $\Pr(c_k = j) = \pi_j$
- $f(\cdot)$  est la forme paramétrique d'un cluster (e.g. gaussienne de moyenne et matrice de covariance inconnues)
- Objectif : Estimer  $\theta = \{U_{1:K}, \pi_{1:K}\}$  étant donné  $\mathbf{z}_{1:n}$
- Approche bayésienne empirique : maximiser  $p(\mathbf{z}_{1:n} | \theta) = \prod_{k=1}^n p(\mathbf{z}_k | \theta) = \prod_{k=1}^n \sum_{j=1}^K \pi_j f(\mathbf{z}_k | U_j)$

# Classification basée sur un modèle

Dempster et al., 1977

- $F$  est supposé être un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$
- $U_1, \dots, U_K$  sont les centroïdes des clusters
- $\pi_1, \dots, \pi_K$  sont les tailles relatives des clusters ( $\sum_{j=1}^K \pi_j = 1$ ) de telle façon que  $\Pr(c_k = j) = \pi_j$
- $f(\cdot)$  est la forme paramétrique d'un cluster (e.g. gaussienne de moyenne et matrice de covariance inconnues)
- Objectif : Estimer  $\theta = \{U_{1:K}, \pi_{1:K}\}$  étant donné  $\mathbf{z}_{1:n}$
- Approche bayésienne empirique : maximiser  $p(\mathbf{z}_{1:n} | \theta) = \prod_{k=1}^n p(\mathbf{z}_k | \theta) = \prod_{k=1}^n \sum_{j=1}^K \pi_j f(\mathbf{z}_k | U_j)$
- Algorithme Expectation-Maximization

# Sommaire

- 1 Introduction
- 2 Modèles bayésiens non paramétriques**
- 3 Algorithme MCMC
- 4 Classification non supervisée dynamique
- 5 Conclusion

# Cadre Bayésien

- Paramètre  $\theta$  et observations  $z$
- Distribution a priori  $p(\theta)$  et vraisemblance  $p(z|\theta)$
- Distribution a posteriori obtenue par le théorème de Bayes

$$p(\theta|z) = \frac{p(\theta)p(z|\theta)}{p(z)}$$

# Markov Chain Monte Carlo

- En général impossible de déterminer analytiquement  $p(\theta|z)$ 
  - ▶ Dimension importante
  - ▶ Pas de forme paramétrique simple (e.g. gaussienne...)
- Développement des méthodes de Monte Carlo
- Comment obtenir des échantillons  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}$  approximativement distribués selon  $p(\theta|z)$  ?
- $\theta = \{\theta_1, \dots, \theta_p\}$

## Echantillonneur de Gibbs

Pour  $i = 1, \dots, N$ ,

- Pour  $k = 1, \dots, p$ , échantillonner  $\theta_k^{(i)}$  selon  $p(\theta_k | \theta_1^{(i)}, \dots, \theta_{k-1}^{(i)}, \theta_{k+1}^{(i-1)}, \dots, \theta_p^{(i-1)}, z)$

# Estimation de densité bayésienne (paramétrique)

- Données  $\mathbf{z}_1, \dots, \mathbf{z}_n$  i.i.d. selon une distribution inconnue  $F$
- $F$  est un mélange de distributions  $F = \sum_{j=1}^K \pi_j f(\cdot | U_j)$
- Cadre bayésien complet : distributions a priori sur les poids et les centroïds

$$\pi_{1:K} \sim \mathcal{D}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right)$$

$$U_j \sim \mathbb{G}_0 \text{ pour } j = 1, \dots, K$$

# Estimation de densité bayésienne (paramétrique)

- Modèle hiérarchique

$$\pi_{1:K} \sim \mathcal{D}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right)$$

$$U_j \sim \mathbb{G}_0 \text{ pour } j = 1, \dots, K$$

Pour  $k = 1, \dots, n$

$$c_k \sim \pi_{1:K}$$

$$\mathbf{z}_k | c_k, U_{1:K} \sim f(\cdot | U_{c_k})$$

- Objectif : Estimer  $p(\pi_{1:K}, U_{1:K}, c_{1:n} | \mathbf{z}_{1:n})$

# Modèle hiérarchique non paramétrique

## Processus de Dirichlet à mélange

- Limite du modèle paramétrique quand  $K \rightarrow \infty$
- Modèle bayésien hiérarchique

$$\pi_{1:K} \sim \mathcal{D}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right)$$
$$U_j \sim \mathbb{G}_0 \text{ pour } j = 1, \dots, K$$

Pour  $k = 1, \dots, n$

$$c_k \sim \pi_{1:K}$$
$$\mathbf{z}_k | c_k, U_{1:K} \sim f(\cdot | U_{c_k})$$

- $F$  prend la forme d'un mélange infini (dénombrable)  
 $F = \sum_{j=1}^{\infty} \pi_j f(\cdot | U_j)$

# Modèle hiérarchique non paramétrique

## Processus de Dirichlet à mélange

- Distribution à deux paramètres sur  $F$
- $\alpha > 0$  règle la taille relative des clusters :
  - ▶ Si  $\alpha \rightarrow 0$ , un cluster de taille  $\simeq 1$
  - ▶ Si  $\alpha \rightarrow \infty$ , infinité de clusters de poids similaires
- $\mathbb{G}_0$  est la distribution a priori sur les centroïdes des clusters
- Non paramétrique : Le nombre d'inconnues (clusters) augmente avec les données

# Simulations

# Simulations

# Simulations

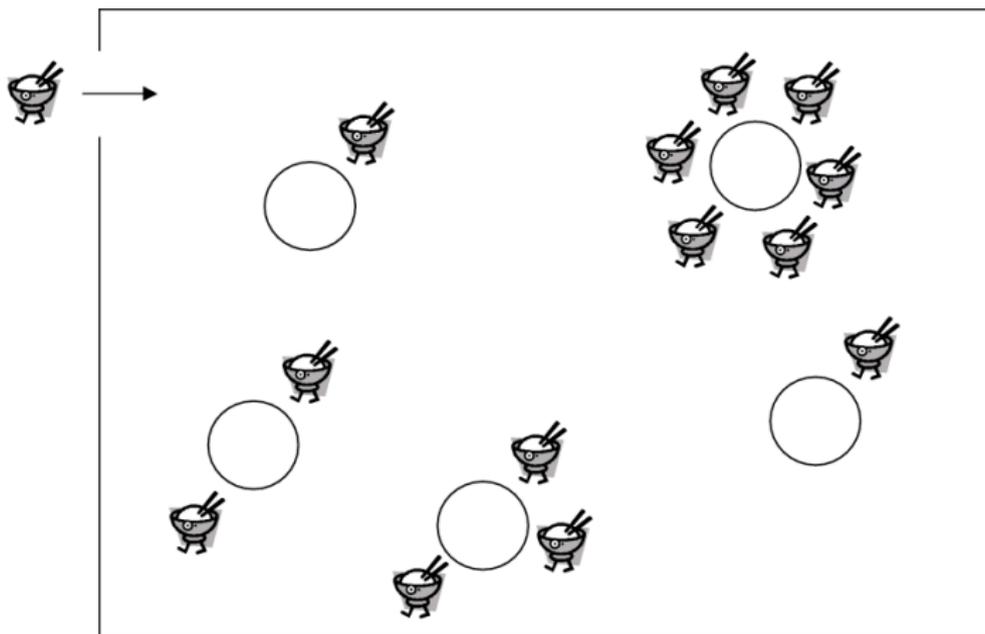
# Sommaire

- 1 Introduction
- 2 Modèles bayésiens non paramétriques
- 3 Algorithme MCMC**
- 4 Classification non supervisée dynamique
- 5 Conclusion

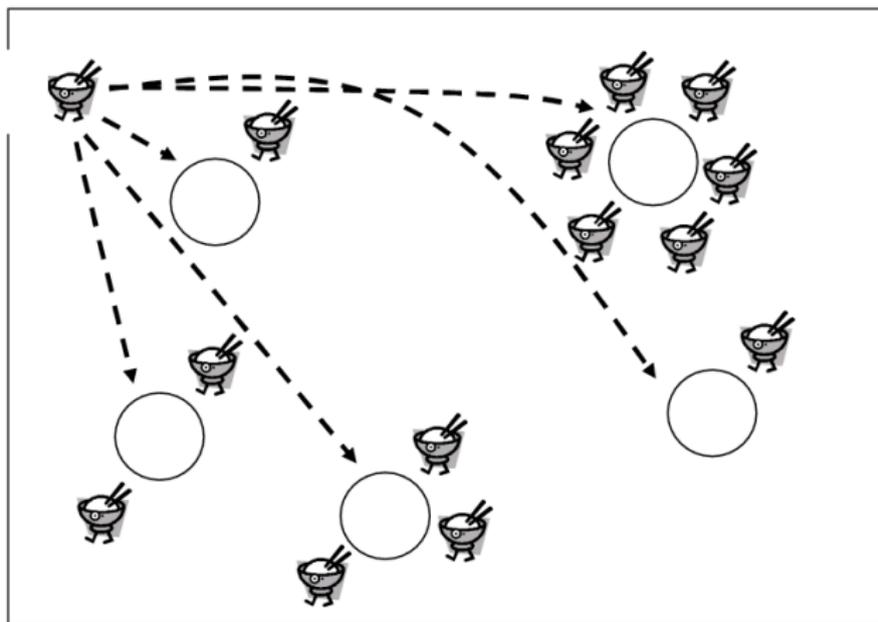
# Processus du restaurant chinois

- Poids  $\pi_{1:\infty}$ , centroïdes des clusters  $U_{1:\infty}$ , variables d'allocation  $c_{1:n}$
- Intégration analytique selon les paramètres (de dimension infinie)  $\pi_{1:\infty}$  et calculer  $\Pr(c_k | c_{1:k-1})$
- processus du restaurant chinois

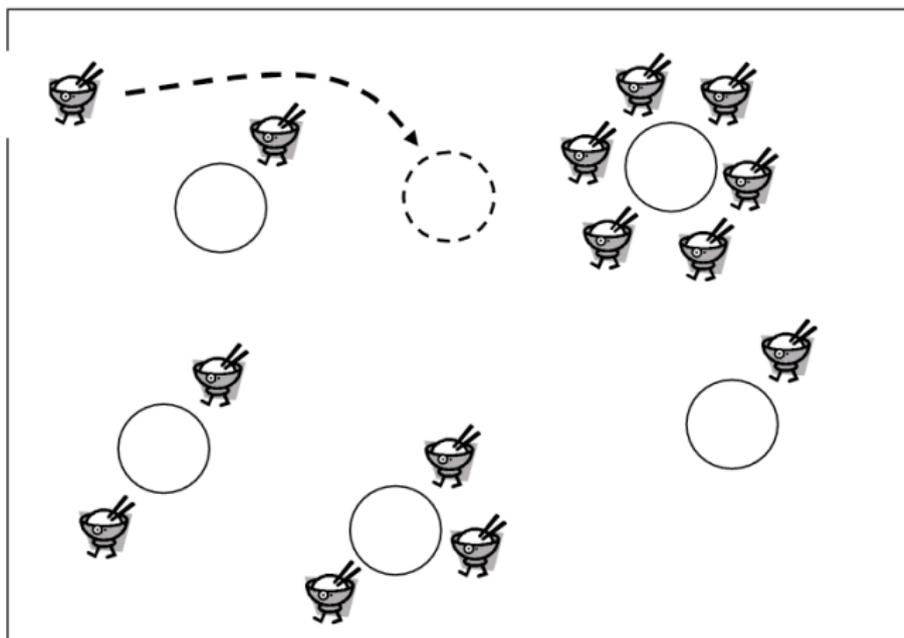
# Processus du restaurant chinois



# Processus du restaurant chinois



# Processus du restaurant chinois



# Processus du restaurant chinois

- Une donnée  $k$  est associée
  - ▶ à un cluster existant avec une probabilité proportionnelle au nombre de données associées à ce cluster
  - ▶ à un nouveau cluster avec une probabilité proportionnelle à  $\alpha$
- Effet de clustering

# Echantillonneur de Gibbs

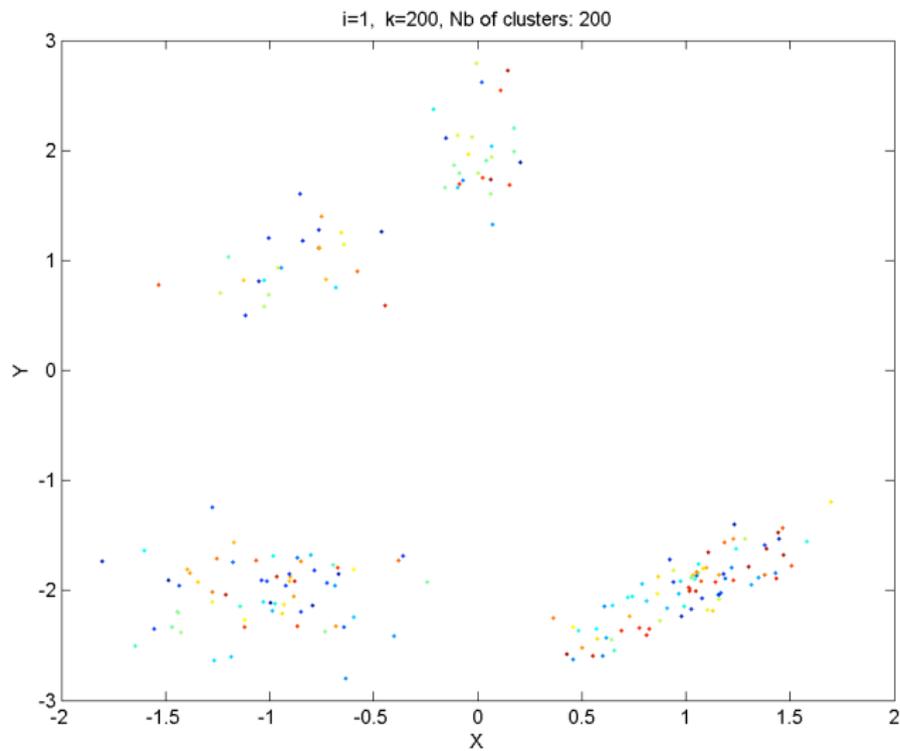
- Sous certaines conditions, on peut également intégrer selon les  $U_j$
- Objectif : Obtenir des échantillons  $c_{1:n}^{(1)}, c_{1:n}^{(2)}, \dots, c_{1:n}^{(N)}$  approximativement distribués selon  $\Pr(c_{1:n} | \mathbf{z}_{1:n})$

## Echantillonneur de Gibbs pour la classification non supervisée

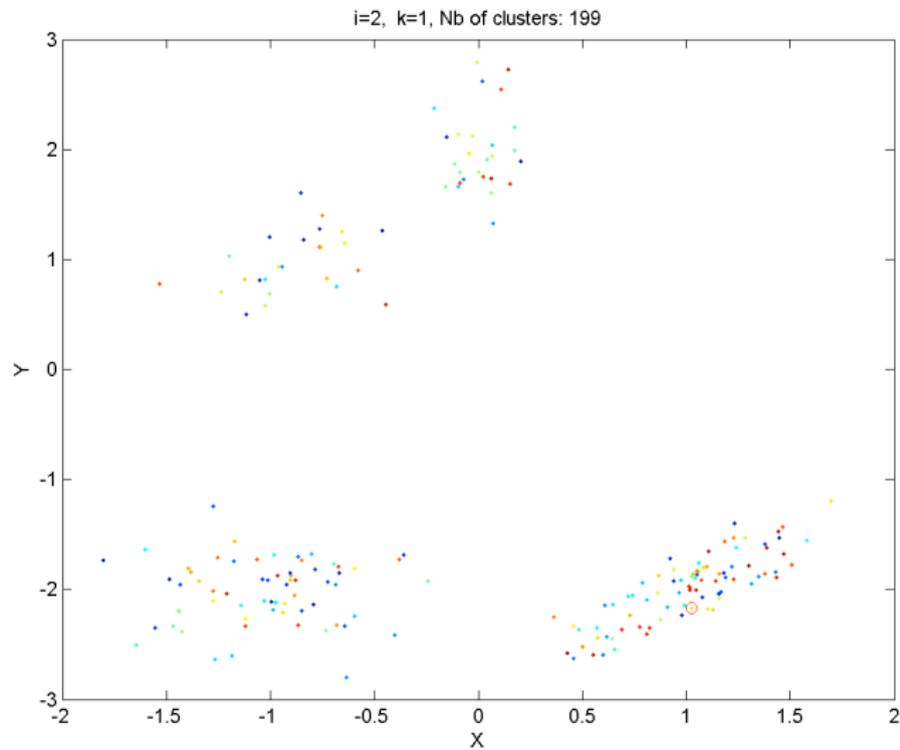
Pour  $i = 1, \dots, N$ ,

- Pour  $k = 1, \dots, n$ , échantillonner  $c_k^{(i)}$  selon  $\Pr(c_k | c_1^{(i)}, \dots, c_{k-1}^{(i)}, c_{k+1}^{(i-1)}, \dots, c_n^{(i-1)}, \mathbf{z}_{1:n})$

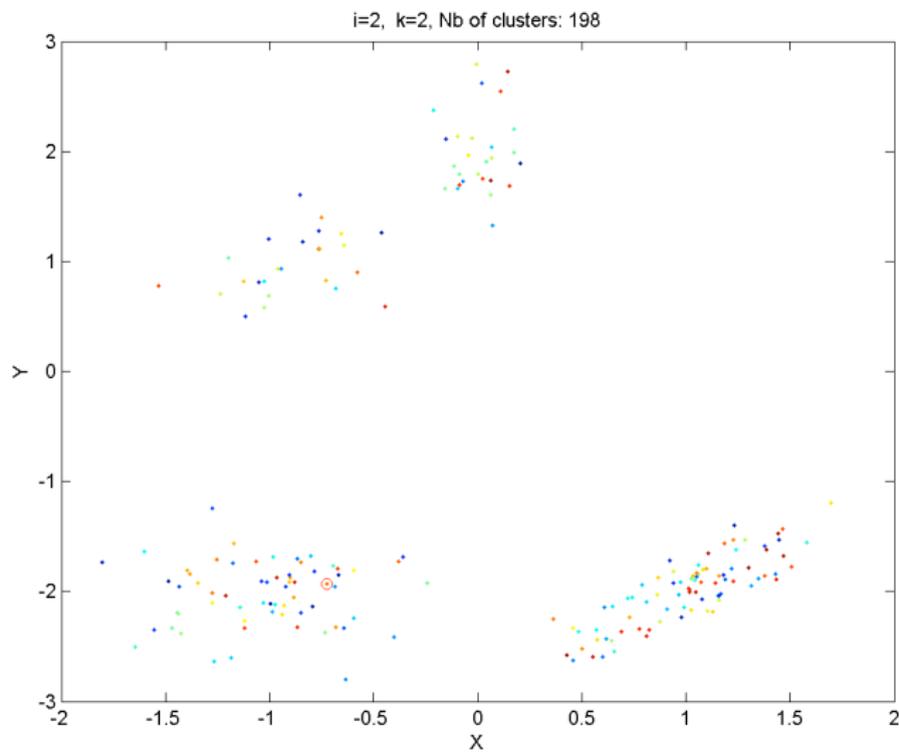
# Echantillonneur de Gibbs



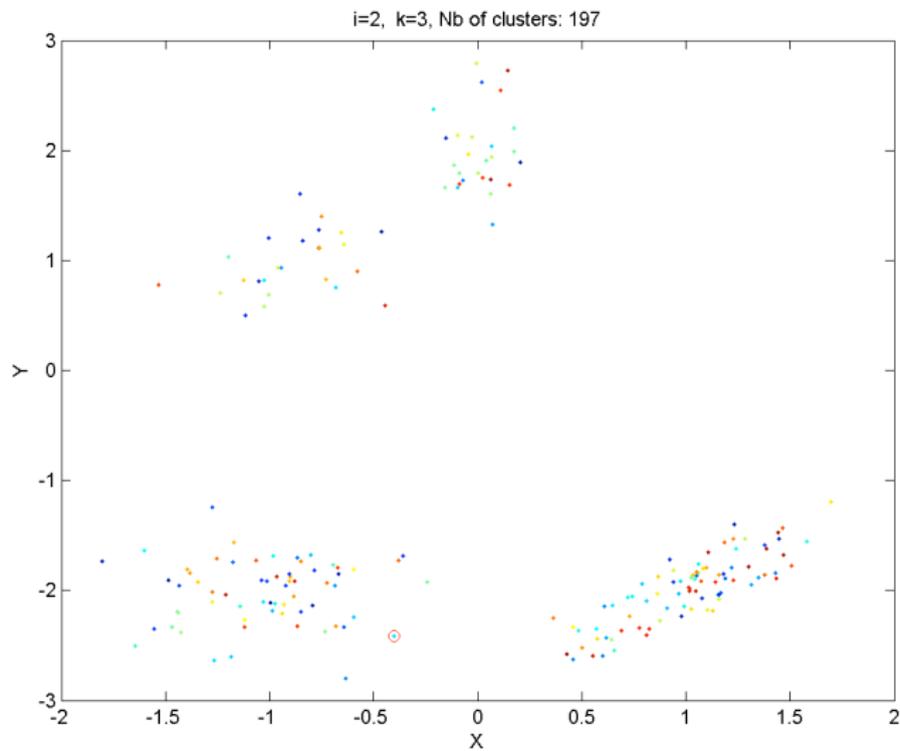
# Echantillonneur de Gibbs



# Echantillonneur de Gibbs



# Echantillonneur de Gibbs



# Echantillonneur de Gibbs

# Méthodes bayésiennes non paramétriques pour la classification non supervisée

## Résumé

- Basé sur un modèle : la forme paramétrique d'un cluster doit être définie (e.g. gaussienne, etc.)
- Approche bayésienne
  - ▶ Distribution a priori sur la taille relative des clusters ( $\alpha$ )
  - ▶ Distribution a priori sur les centroïdes des clusters ( $\mathbb{G}_0$ )
- Nombre de clusters inconnu et estimé à partir des données
- Algorithmes de simulation efficaces
- Bruit de fond/autres paramètres inconnus peuvent être pris en compte

# Sommaire

- 1 Introduction
- 2 Modèles bayésiens non paramétriques
- 3 Algorithme MCMC
- 4 Classification non supervisée dynamique**
- 5 Conclusion

# Classification non supervisée dynamique

- Clustering statique : hypothèse d'échangeabilité des données  $\mathbf{z}_1, \dots, \mathbf{z}_n$
- Si les données sont obtenues séquentiellement, cette hypothèse peut être trop forte
- Nombre de clusters, les centroïdes/tailles des clusters peuvent évoluer en fonction du temps

# Clustering dynamique

- Ensemble de données  $\mathbf{z}_{k,n}$   $k = 1, \dots, n$ ,  $t = 1, \dots, T$

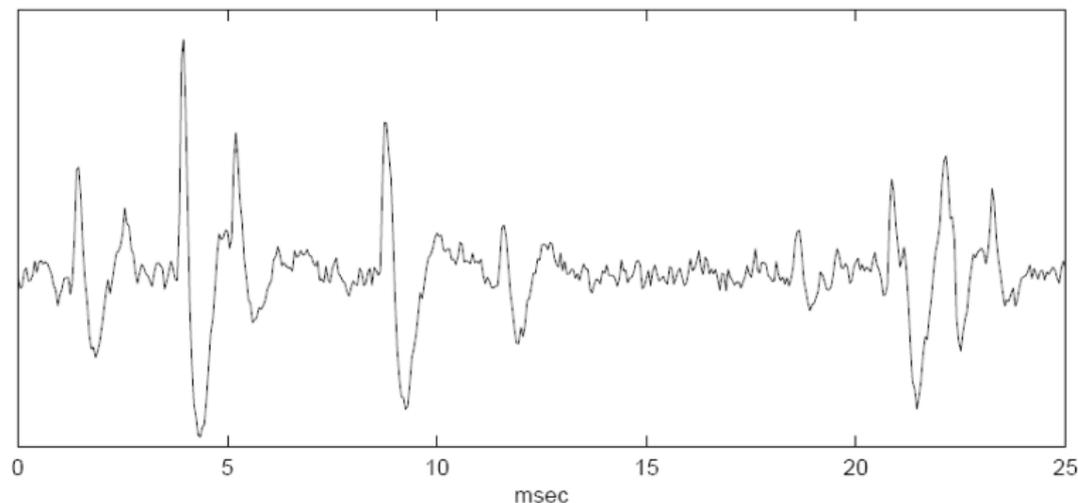
# Clustering dynamique

- Ensemble de données  $\mathbf{z}_{k,n}$   $k = 1, \dots, n$ ,  $t = 1, \dots, T$

# Exemple : Classification de pics neuronaux

Bar-Hillel et al., J. of Neuroscience Methods, 2006

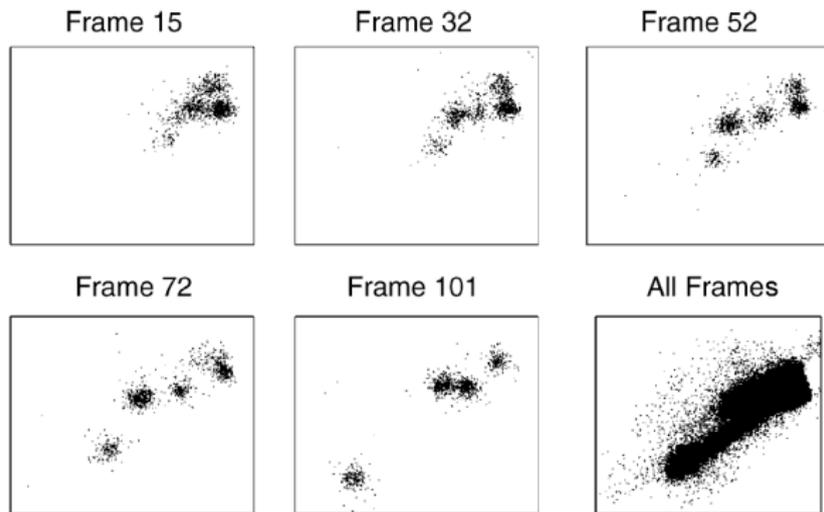
- Brefs potentiels d'action enregistrés par une micro-électrode
- Objectif : classer les signaux afin d'attribuer chaque potentiel à un neurone particulier
- Dérive de l'électrode, évolution des formes des pics et du bruit de fond...



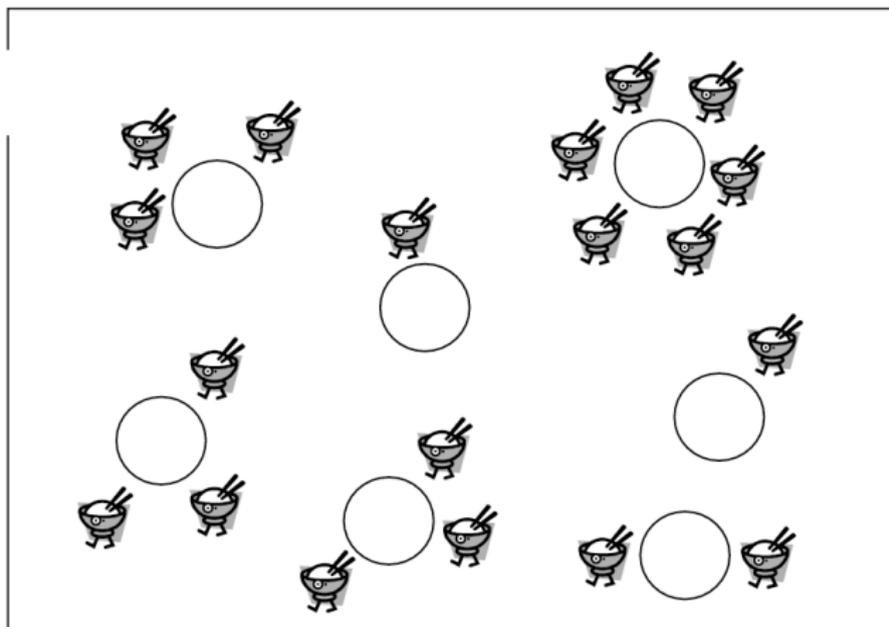
# Exemple : Classification de pics neuronaux

Bar-Hillel et al., J. of Neuroscience Methods, 2006

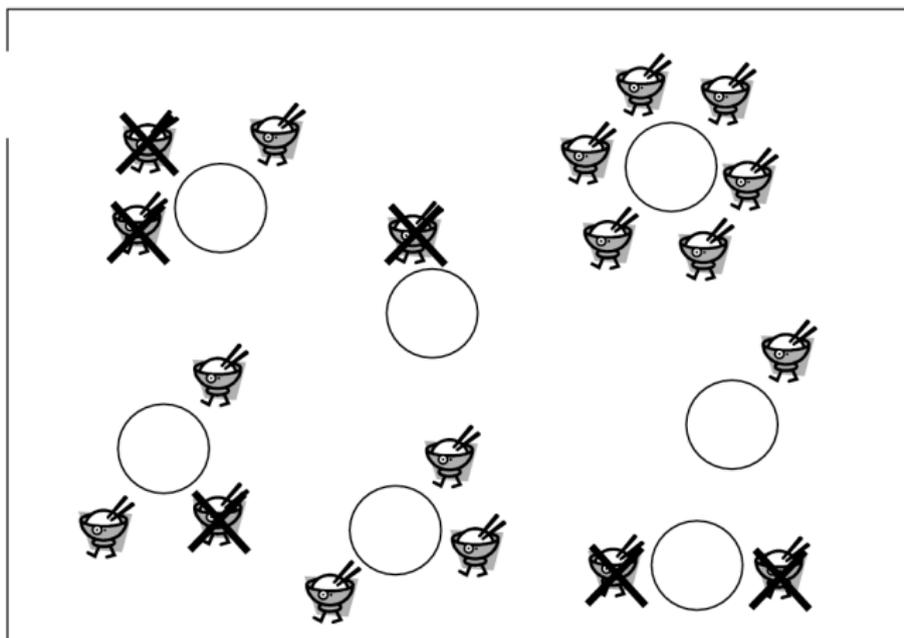
- Brefs potentiels d'action enregistrés par une micro-électrode
- Objectif : classer les signaux afin d'attribuer chaque potentiel à un neurone particulier
- Dérive de l'électrode, évolution des formes des pics et du bruit de fond...



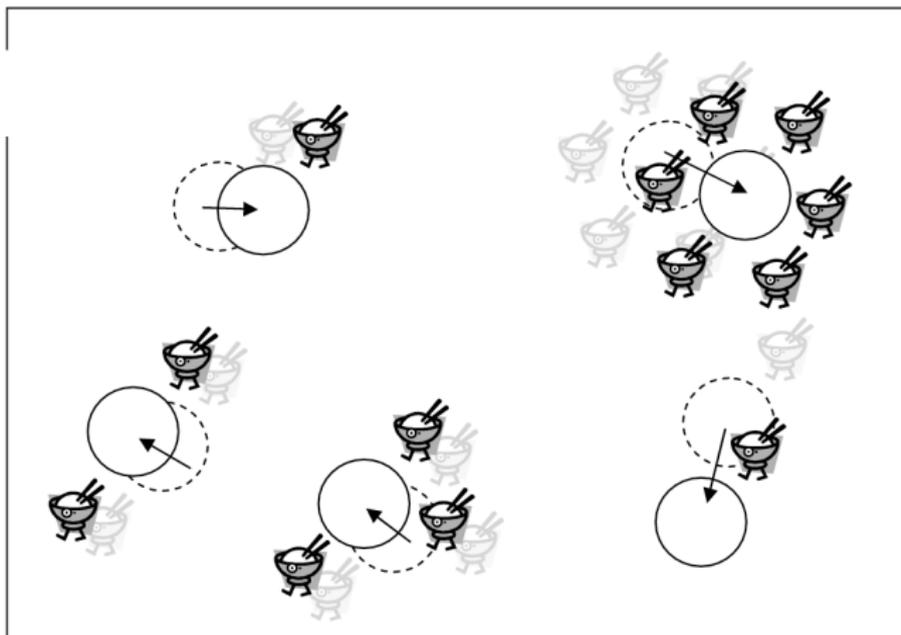
# Processus du restaurant chinois dynamique



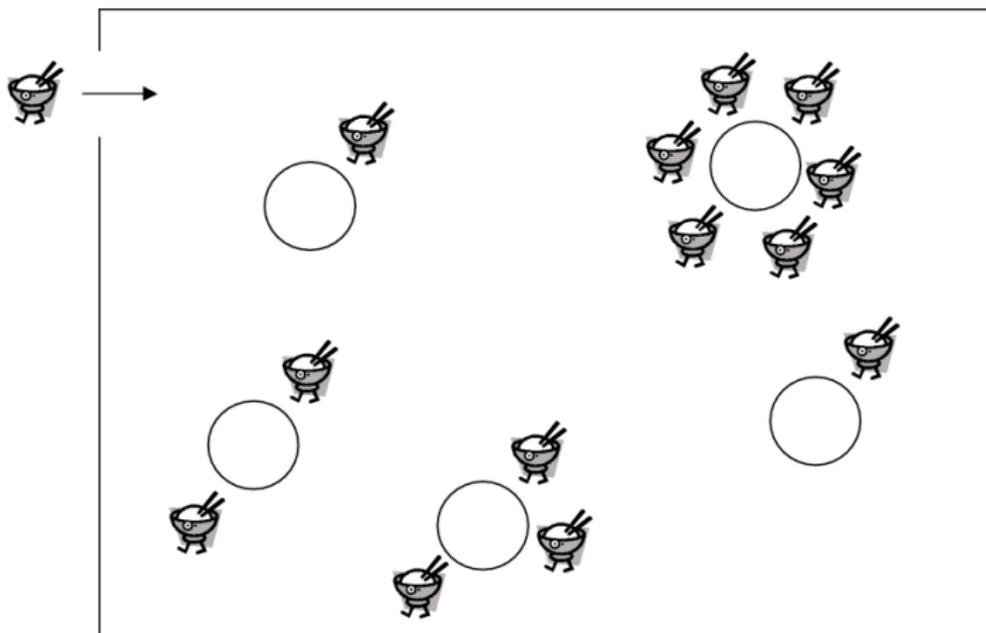
# Processus du restaurant chinois dynamique



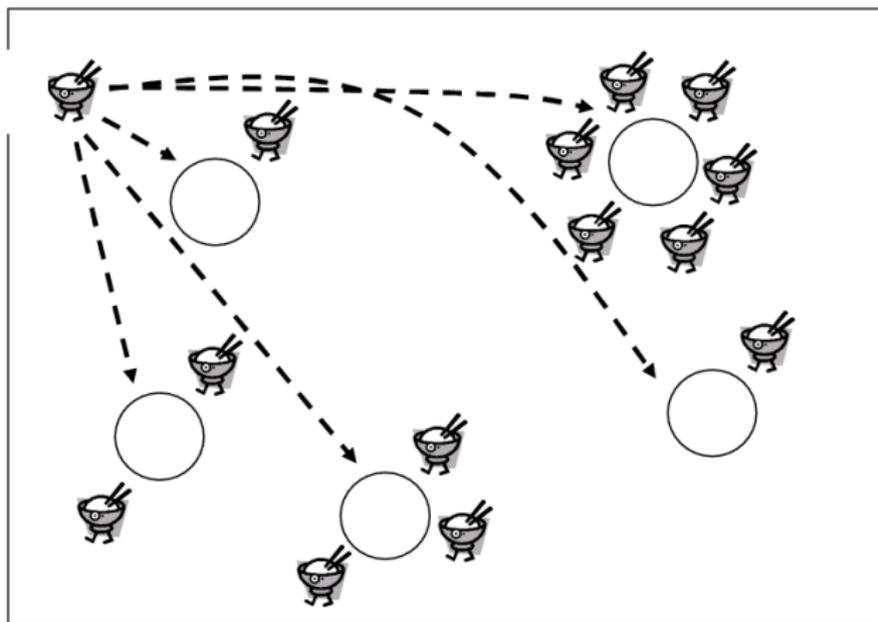
# Processus du restaurant chinois dynamique



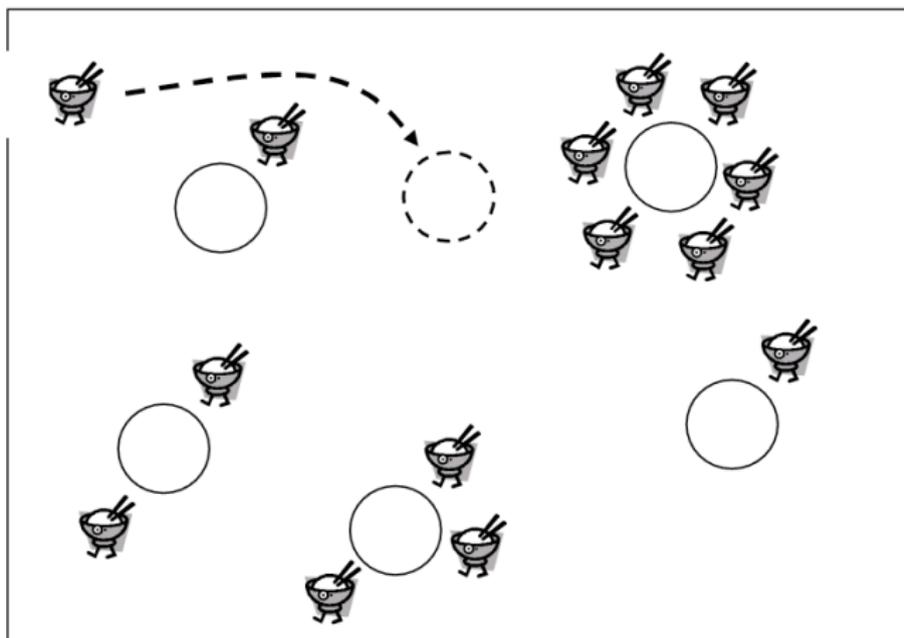
# Processus du restaurant chinois dynamique



# Processus du restaurant chinois dynamique



# Processus du restaurant chinois dynamique



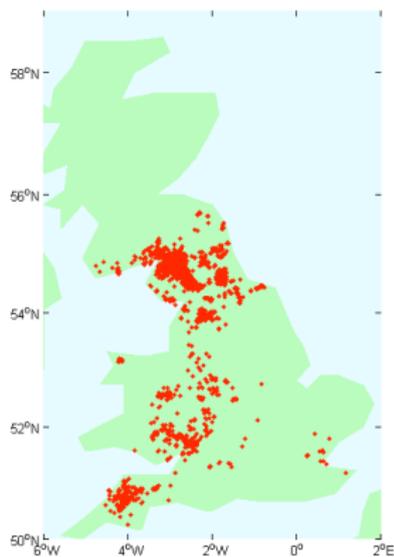
# Algorithmes

- On peut dans certains cas intégrer également les centroïdes des clusters
- Échantillons pour approcher  $\Pr(c_{1:n,1:T} | z_{1:n,1:T})$
- Algorithmes de Monte Carlo efficaces hors et en ligne

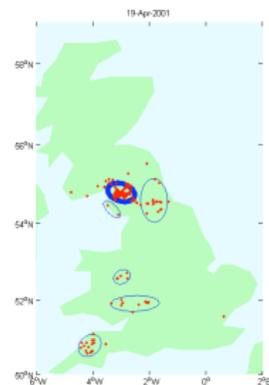
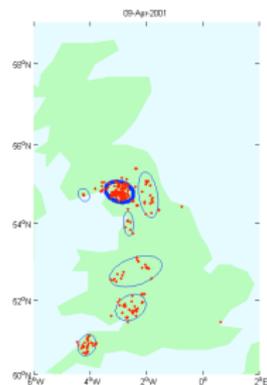
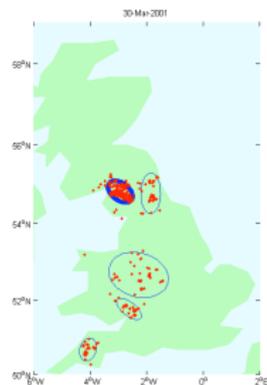
# Simulation

- Données simulées avec bruit de fond

# Maladie de la fièvre aphteuse



# Maladie de la fièvre aphteuse



# Sommaire

- 1 Introduction
- 2 Modèles bayésiens non paramétriques
- 3 Algorithme MCMC
- 4 Classification non supervisée dynamique
- 5 Conclusion**

# Conclusion

Méthodes bayésiennes non paramétriques : Processus de Dirichlet

- Basé sur un modèle : forme paramétrique d'un cluster

# Conclusion

Méthodes bayésiennes non paramétriques : Processus de Dirichlet

- Basé sur un modèle : forme paramétrique d'un cluster
- Bayésien : distribution a priori sur les centroïds/taille des clusters

# Conclusion

## Méthodes bayésiennes non paramétriques : Processus de Dirichlet

- Basé sur un modèle : forme paramétrique d'un cluster
- Bayésien : distribution a priori sur les centroïds/taille des clusters
- Non paramétrique : Nombre de clusters est directement estimé à partir des données

# Conclusion

## Méthodes bayésiennes non paramétriques : Processus de Dirichlet

- Basé sur un modèle : forme paramétrique d'un cluster
- Bayésien : distribution a priori sur les centroïds/taille des clusters
- Non paramétrique : Nombre de clusters est directement estimé à partir des données
- Algorithmes de Monte Carlo efficaces pour estimer ces modèles

# Conclusion

## Méthodes bayésiennes non paramétriques : Processus de Dirichlet

- Basé sur un modèle : forme paramétrique d'un cluster
- Bayésien : distribution a priori sur les centroïds/taille des clusters
- Non paramétrique : Nombre de clusters est directement estimé à partir des données
- Algorithmes de Monte Carlo efficaces pour estimer ces modèles
- Versions statiques/dynamiques

# Bibliographie



F. Caron, M. Davy, A. Doucet

Generalized Polya Urn for time-varying Dirichlet Process Mixtures  
*International Conference on Uncertainty in Artificial Intelligence*, 2007.



A.P. Dempster, N.M. Laird and D.B. Rubin

Maximum Likelihood from Incomplete Data via the EM algorithm  
*Journal of the Royal Statistical Society*, 1977.



T. Ferguson

A Bayesian analysis of some nonparametric problems,  
*The annals of statistics*, vol. 1, pp. 209-230, 1973.



R.M. Neal

Markov chain sampling methods for Dirichlet process mixture models,  
*Journal of Computational and Graphical Statistics*, 9 :249-265, 2000



C. Robert et G. Casella

Monte Carlo statistical Methods,  
*Springer-Verlag*, 1999.

The end

**Merci pour votre attention**