

MCMC et approximations en champ moyen pour les modèles de Markov

Gersende FORT

LTCI
CNRS - TELECOM ParisTech

En collaboration avec Florence FORBES (Projet MISTIS, INRIA Rhône-Alpes).

Basé sur l'article: *A convergence theorem for Variational EM-like algorithms: application to image segmentation*. IEEE Transactions on Image Processing, 16(3):824-837,2007

- ▶ y : Vecteur d'observations, de (grande) dimension N .
- ▶ ψ : paramètre qui gouverne la loi des observations $\mathcal{L}_Y(\cdot; \psi)$

↔ **Objectif**: estimer le paramètre ψ au sens du maximum de vraisemblance

$$\max_{\psi} \mathcal{L}_Y(y; \psi) ,$$

dans un contexte de données latentes, à structure de dépendance complexe.

1. Inférence dans les modèles à données manquantes
 - ▶ EM
 - ▶ Champ de markov caché
2. EM variationnel
 - ▶ VEM
 - ▶ Champ moyen
3. Coupler approximation type champ moyen, et simulation
 - ▶ Approches purement déterministes
 - ▶ Approches purement stochastiques
 - ▶ Approches déterministes + stochastiques.
4. Comparaison des approches: estimation de paramètres dans un modèle de Potts caché.

I. Inférence dans les modèles à données manquantes

Algorithme EM

- ▶ Résolution itérative du problème d'optimisation $\max_{\psi} \mathcal{L}_Y(y; \psi)$.
- ▶ Chaque itération comporte deux étapes:

(i) Etape E :

$$\begin{aligned} Q(\psi; \psi^{(t)}) &:= \int \ln \mathcal{L}_{(Z,Y)}(z,y; \psi) \mathcal{L}_{Z|Y}(z|y; \psi^{(t)}) \\ &= \mathbb{E} \left[\ln \mathcal{L}_{(Z,Y)}(Z,y; \psi) | y; \psi^{(t)} \right] . \end{aligned}$$

(ii) Etape M :

$$\psi^{(t+1)} := \max_{\psi} Q(\psi; \psi^{(t)}) .$$

Algorithme EM

- ▶ Résolution itérative du problème d'optimisation $\max_{\psi} \mathcal{L}_Y(y; \psi)$.
- ▶ Chaque itération comporte deux étapes:

(i) Etape E :

$$\begin{aligned} Q(\psi; \psi^{(t)}) &:= \int \ln \mathcal{L}_{(Z,Y)}(z, y; \psi) \mathcal{L}_{Z|Y}(z|y; \psi^{(t)}) \\ &= \mathbb{E} \left[\ln \mathcal{L}_{(Z,Y)}(Z, y; \psi) | y; \psi^{(t)} \right] . \end{aligned}$$

(ii) Etape M :

$$\psi^{(t+1)} := \max_{\psi} Q(\psi; \psi^{(t)}) .$$

- ▶ Par construction :

$$\mathcal{L}_Y(y; \psi^{(t+1)}) \geq \mathcal{L}_Y(y; \psi^{(t)}) .$$

- ▶ Convergence de la suite $\{\psi^{(t)}\}$ vers un max local de $\psi \mapsto \mathcal{L}_Y(y; \psi)$.

Hidden Markov Random Field (discret):

- ▶ loi des données latentes: Champ de Markov (discret)

$$\mathcal{L}_Z(z; \psi) = W(\beta)^{-1} \exp(-H(z; \beta)), \quad \psi = (\beta, \cdot) \quad z \in \mathcal{Z}$$

avec

$$W(\beta) = \sum_z \exp(-H(z; \beta)) \quad H(z; \beta) = \sum_{c \in \mathcal{C}} V_c(z_c) .$$

Hidden Markov Random Field (discret):

- ▶ loi des données latentes: Champ de Markov (discret)

$$\mathcal{L}_Z(z; \psi) = W(\beta)^{-1} \exp(-H(z; \beta)) , \quad \psi = (\beta, \cdot) \quad z \in \mathcal{Z}$$

avec

$$W(\beta) = \sum_z \exp(-H(z; \beta)) \quad H(z; \beta) = \sum_{c \in \mathcal{C}} V_c(z_c) .$$

- ▶ Attache aux données :

$$\mathcal{L}_{Y|Z}(y|z; \psi) = \prod_{k=1}^N p_k(y_k | z_k; \theta) \quad \psi = (\beta, \theta) .$$

Hidden Markov Random Field (discret):

- ▶ loi des données latentes: Champ de Markov (discret)

$$\mathcal{L}_Z(z; \psi) = W(\beta)^{-1} \exp(-H(z; \beta)) , \quad \psi = (\beta, \cdot) \quad z \in \mathcal{Z}$$

avec

$$W(\beta) = \sum_z \exp(-H(z; \beta)) \quad H(z; \beta) = \sum_{c \in \mathcal{C}} V_c(z_c) .$$

- ▶ Attache aux données :

$$\mathcal{L}_{Y|Z}(y|z; \psi) = \prod_{k=1}^N p_k(y_k | z_k; \theta) \quad \psi = (\beta, \theta) .$$

- ▶ Difficultés

- ▶ Loi sur un espace de grande dimension.
- ▶ Constante de normalisation $W(\beta)$ incalculable.

EM pour inférence dans HMRF non implémentable car

► Etape E :

- intégration sous une loi complexe $\mathcal{L}_{Z|Y}$,
- en grande dimension: ex. $\mathcal{Z} = \{1, \dots, K\}^N$

EM pour inférence dans HMRF non implémentable car

▶ Etape E :

- ▶ intégration sous une loi complexe $\mathcal{L}_{Z|Y}$,
- ▶ en grande dimension: ex. $\mathcal{Z} = \{1, \dots, K\}^N$

▶ Etape M :

- ▶ Il faut savoir maximiser en $\psi = (\beta, \theta)$

$$\sum_z \mathcal{L}_{Y|Z}(y|z; \theta) \mathcal{L}_{Z|Y}(z|y; \psi^{(t)}) - \ln W(\beta) - \sum_z H(z; \beta) \mathcal{L}_{Z|Y}(z|y; \psi^{(t)})$$

- ▶ En pratique, la maximisation en θ : explicite.
- ▶ Mais la maximisation en β demande la connaissance de $W(\beta)$.

II. Variational EM et Approximation en champ moyen

Autre lecture de l'EM:

- Pour toute probabilité q sur \mathcal{Z} ,

$$\begin{aligned} \ln \mathcal{L}_Y(y; \psi) &= \int_{\mathcal{Z}} \mathcal{L}_{(Z,Y)}(z,y; \psi) dz = \ln \int_{\mathcal{Z}} \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} q(z) dz , \\ &\geq \int_{\mathcal{Z}} \ln \left\{ \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} \right\} q(z) dz =: F(q, \psi) . \end{aligned}$$

Autre lecture de l'EM:

- Pour toute probabilité q sur \mathcal{Z} ,

$$\begin{aligned}\ln \mathcal{L}_Y(y; \psi) &= \int_{\mathcal{Z}} \mathcal{L}_{(Z,Y)}(z,y; \psi) dz = \ln \int_{\mathcal{Z}} \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} q(z) dz , \\ &\geq \int_{\mathcal{Z}} \ln \left\{ \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} \right\} q(z) dz =: F(q, \psi) .\end{aligned}$$

Avec égalité : $\ln \mathcal{L}_Y(y; \psi) = F(\mathcal{L}_{Z|Y}(z|y; \psi) , \psi)$

Autre lecture de l'EM:

- Pour toute probabilité q sur \mathcal{Z} ,

$$\begin{aligned} \ln \mathcal{L}_Y(y; \psi) &= \int_{\mathcal{Z}} \mathcal{L}_{(Z,Y)}(z,y; \psi) dz = \ln \int_{\mathcal{Z}} \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} q(z) dz , \\ &\geq \int_{\mathcal{Z}} \ln \left\{ \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} \right\} q(z) dz =: F(q, \psi) . \end{aligned}$$

Avec égalité: $\ln \mathcal{L}_Y(y; \psi) = F(\mathcal{L}_{Z|Y}(z|y; \psi), \psi)$

- EM = maximisation alternée de F : pour $(q^{(t)}, \psi^{(t)})$
 - mise à jour de la loi:

$$q^{(t+1)} = \operatorname{argmax}_q F(q, \psi^{(t)}) \quad \text{rép.:} \quad q^{(t+1)} = \mathcal{L}_{Z|Y}(z|y; \psi^{(t)})$$

Autre lecture de l'EM:

- Pour toute probabilité q sur \mathcal{Z} ,

$$\begin{aligned} \ln \mathcal{L}_Y(y; \psi) &= \int_{\mathcal{Z}} \mathcal{L}_{(Z,Y)}(z,y; \psi) dz = \ln \int_{\mathcal{Z}} \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} q(z) dz , \\ &\geq \int_{\mathcal{Z}} \ln \left\{ \frac{\mathcal{L}_{(Z,Y)}(z,y; \psi)}{q(z)} \right\} q(z) dz =: F(q, \psi) . \end{aligned}$$

Avec égalité: $\ln \mathcal{L}_Y(y; \psi) = F(\mathcal{L}_{Z|Y}(z|y; \psi), \psi)$

- EM = maximisation alternée de F : pour $(q^{(t)}, \psi^{(t)})$
 - mise à jour de la loi:

$$q^{(t+1)} = \operatorname{argmax}_q F(q, \psi^{(t)}) \quad \text{rép.:} \quad q^{(t+1)} = \mathcal{L}_{Z|Y}(z|y; \psi^{(t)})$$

- mise à jour du paramètre:

$$\psi^{(t+1)} = \operatorname{argmax}_{\psi} F(q^{(t)}, \psi) = \operatorname{argmax}_{\psi} Q(\psi, \psi^{(t)}) .$$

Variational EM

- ▶ Relâcher les contraintes d'optimisation de l'EM :
 - ▶ considérer un sous-espace des probabilités sur \mathcal{Z} , celles de forme produit

$$q(z) = \prod_{k=1}^N q_k(z_k) .$$

- ▶ ne rien changer pour ψ .

Variational EM

- ▶ Relâcher les contraintes d'optimisation de l'EM :
 - ▶ considérer un sous-espace des probabilités sur \mathcal{Z} , celles de forme produit

$$q(z) = \prod_{k=1}^N q_k(z_k) .$$

- ▶ ne rien changer pour ψ .
- ▶ Choisir q revient à résoudre l'équation au point fixe

$$\ln q_k(e) = c_k + \sum_z \ln \mathcal{L}_{Z|Y}(z|y; \psi^{(t)}) \left\{ \delta_e(z_k) \prod_{j \neq k} q_j(z_j) \right\} ,$$

qui ne dépend que des $q_j, j \in \mathcal{N}_k$, du fait de la structure markovienne.

Variational EM pour inférence dans les HMRF

- ▶ Résultats de convergence : oui.
- ▶ Identification des points limites: ...
- ▶ Pour l'inférence dans les HMRF
 - ▶ permet de contourner une des difficultés rencontrée par l'EM: étape E.
 - ▶ mais la mise à jour de ψ demande toujours le calcul de $W(\beta)$.

Champ moyen

- ▶ A l'origine, pour approcher le calcul de l'espérance sous une loi de Gibbs.
- ▶ Négliger les fluctuations des voisins, en les fixant à leur valeur moyenne **pour la loi ainsi définie**.

Champ moyen

- ▶ A l'origine, pour approcher le calcul de l'espérance sous une loi de Gibbs.
- ▶ Négliger les fluctuations des voisins, en les fixant à leur valeur moyenne **pour la loi ainsi définie**.
- ▶ En pratique:
 - ▶ définir une loi produit $q(z) = \prod_{k=1}^N q_k(z_k)$,
 - ▶ solution d'une équation au point fixe. Dans le contexte :

mesure de Gibbs \longleftrightarrow loi a posteriori $\mathcal{L}_{Z|Y}(z|y; \psi^{(t)})$

même équation au point fixe que pour VEM.

Champ moyen

- ▶ A l'origine, pour approcher le calcul de l'espérance sous une loi de Gibbs.
- ▶ Négliger les fluctuations des voisins, en les fixant à leur valeur moyenne **pour la loi ainsi définie**.
- ▶ En pratique:
 - ▶ définir une loi produit $q(z) = \prod_{k=1}^N q_k(z_k)$,
 - ▶ solution d'une équation au point fixe. Dans le contexte :

mesure de Gibbs \longleftrightarrow loi a posteriori $\mathcal{L}_{Z|Y}(z|y; \psi^{(t)})$

même équation au point fixe que pour VEM.

- ▶ En conclusion :

Approximation champ moyen \longleftrightarrow Variational EM

III. Approximations déterministes, Approximations par simulation

Approx. purement déterministes

- ▶ ex. Algorithme *Mean-Field*
 - ▶ Etape E : substituée par une espérance sous la loi produit $q^{(t+1)}$
 - ▶ Etape M : propager cette approximation à l'expression de la loi jointe: l'étape de maximisation devient explicite.
- ▶ Pas de résultats sur le comportement asymptotique de l'algorithme.

Approx. purement stochastiques

- ▶ ex. Algorithme *MC2-EM*
 - ▶ Etape E : calcul des espérances par Monte Carlo (MCMC).
 - ▶ Etape M : calcul de la constante de normalisation par Monte Carlo

$$W(\beta) = \sum_z \exp(-H(z; \beta)) = \sum_z \frac{\exp(-H(z; \beta))}{\pi(z)} \pi(z) .$$

- ▶ Dès lors que l'erreur d'approximation due à la simulation est contrôlée : convergence vers les mêmes points limites que l'EM.

VEM et Simulations (I)

- ▶ ex. Algorithme *MC VEM*
 - ▶ Etape E : substituée par une espérance sous la loi produit $q^{(t+1)}$
 - ▶ Etape M : calcul de la constante de normalisation par Monte Carlo

$$W(\beta) = \sum_z \exp(-H(z; \beta)) = \sum_z \frac{\exp(-H(z; \beta))}{\pi(z)} \pi(z) .$$

- ▶ Dès lors que l'erreur d'approximation due à la simulation est contrôlée : convergence vers les mêmes points limites que le VEM.

VEM et Simulations (II)

- ▶ ex. Algorithme *Simulated Field*
 - ▶ Etape E : type VEM en fixant les voisins à une valeur simulée (ne pas résoudre l'équation au point fixe)
 - ▶ Etape M : propager l'approximation champ moyen à l'expression de la loi jointe; l'étape de maximisation devient réalisable.
- ▶ Pas de résultats sur le comportement asymptotique de l'algorithme. A priori, $\{\psi^{(t)}\}$ chaîne de Markov, pas de raisons d'observer une convergence trajectorielle.

IV. Applications

Modèle

- ▶ Champs de Markov (caché): Potts à K classes, voisinage 1er ordre

$$\mathcal{L}_Z(z; \psi) \propto \exp(-\beta \sum_{i \sim j} \langle z_i, z_j \rangle) \quad z_i \in \{e_1, \dots, e_K\} .$$

- ▶ Attache aux données: gaussien

$$\mathcal{L}_{Y|Z}(y|z; \psi) = \prod_{k=1}^N \mathcal{N}(\mu_{z_k}, \sigma_{z_k}^2)(y_k) .$$

- ▶ Paramètre:

$$\beta, \mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2 .$$

- ▶ Possibilité aussi : d'estimer les données cachées (ex. par MAP).

Data 1 : données simulées

- ▶ $K = 2$, images 100×100 .
- ▶ Paramètre θ (attache aux données) : toujours bien estimé.
- ▶ Résultat d'estimation de β et erreur de classification:

algorithm	β	error rate	ref.
true values	0.78	-	-
ind-EM	-	15.91 (0.33)	15.85 (0.26)
Mean Field	0.94 (2.83 10e-2)	10.28 (0.49)	9.77 (0.42)
Simulated Field	0.78 (2.24 10e-2)	10.96 (0.43)	11.04 (0.48)
MCVEM	0.73 (1.77 10e-2)	9.87 (0.42)	9.77 (0.42)
MC2-EM	0.77 (1.44 10e-2)	9.81 (0.39)	9.81 (0.39)
Gibbsian-EM	0.77 (2.23 10e-2)	9.79 (0.40)	9.81 (0.39)

Data 2: données simulées

- ▶ $K = 3$, images 100×100 .
- ▶ Paramètre θ (attache aux données): toujours bien estimé.
- ▶ Résultat d'estimation de β et erreur de classification:

algorithm	β	error rate	ref.
true values	0.90	-	-
ind-EM	-	21.31 (0.60)	21.14 (0.50)
Mean Field	1.03 (2.45 10e-2)	14.03 (0.60)	13.78 (0.59)
Simulated Field	0.90 (2.45 10e-2)	15.67 (0.56)	15.69 (0.64)
MCVEM	0.85 (1.89 10e-2)	14.02 (0.59)	13.78 (0.59)
MC2-EM	0.89 (1.36 10e-2)	13.77 (0.53)	13.79 (0.54)
Gibbsian-EM	0.89 (2.23 10e-2)	13.77 (0.53)	13.79 (0.54)

Data 3: données simulées

- ▶ $K = 4$, images 100×100 .
- ▶ Paramètre θ (attache aux données): toujours bien estimé.
- ▶ Résultat d'estimation de β et erreur de classification:

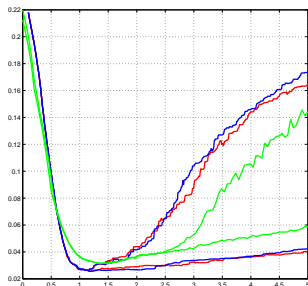
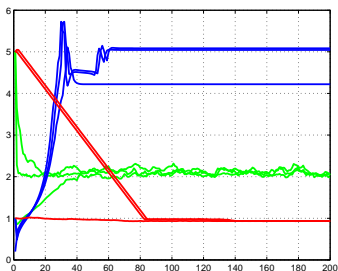
algorithm	β	error rate	ref.
true values	1	-	-
ind-EM	-	24.23 (0.54)	23.87 (0.45)
Mean Field	1.05 (1.95 10e-2)	18.32 (0.51)	18.38 (0.45)
Simulated Field	0.90 (1.64 10e-2)	20.73 (0.55)	20.82 (0.48)
MCVEM	0.81 (1.17 10e-2)	18.66 (0.50)	18.38 (0.45)
MC2-EM	0.89 (1.07 10e-2)	18.15 (0.49)	18.24 (0.47)
Gibbsian-EM	0.89 (1.67 10e-2)	18.14 (0.50)	18.24 (0.47)

Data 4: Logo

- ▶ Image 133×142 ; 2 couleurs.
- ▶ Estimation des paramètres

algorithm	β	μ_1	μ_2	σ_1	σ_2	error rate
true values	-	51	255	130	300	-
ind-EM	-	52	255	128	304	22.69
Mean Field	4.22	53	260	130	306	3.10
Simulated Field	2.15	52	250	128	302	3.42
MCVEM	0.93	50	262	125	305	2.89
MC2-EM	0.91	50	261	125	305	2.89
Gibbsian-EM	1.82	52	251	128	303	2.92

Data 4: Logo (suite)



- ▶ (gauche) Sensibilité aux paramètres d'implémentation *Mean Field*, *Simulated Field* and *MCVEM*
- ▶ (droite) Evolution de l'erreur de classification en fonction de β (et sensibilité à la non-unicité de la solution de l'équation au point fixe)

Data 4: Logo (suite)

- ▶ Image 133×142 ; 2 couleurs.
- ▶ Segmentation :

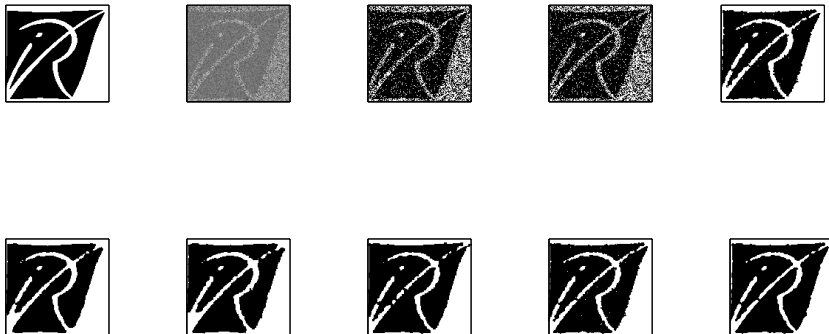


FIG.: [top, from left to right] original image, noise-corrupted image, initial segmentation using kmeans, ind-EM, MC2-EM; [bottom, from left to right] Gibbsian-EM, Simulated Field, Mean Field, MCVEM, MCVEM + Median Filter

Data 5: Image satellite

- ▶ Image 256×256
- ▶ Segmentation :

