

Le peu que l'on sait d'un point de vue théorique sur la qualité du champ moyen

Nathalie Peyrard, dept. MIA - INRA - Toulouse

Florence Forbes, MISTIS INRIA Rhône-Alpes

Journées du réseau MSTGA - 20 et 21 sept. 2007

I. Rappels sur que l'on appelle champ moyen

But des méthodes variationnelles → approximation d'une loi jointe complexe $P(x_1, x_2, \dots, x_n)$ en deux étapes :

1. choix d'une famille \mathcal{Q} "manipulable" de distributions sur (x_1, x_2, \dots, x_n) .
2. "projection" de P sur \mathcal{Q} au sens de la divergence de Kullback-Leibler.

$$\hat{Q} = \arg \min_{Q \in \mathcal{Q}} KL(Q | P)$$

avec

$$KL(Q | P) = \sum_{x_1, x_2, \dots, x_n} Q \ln \frac{P}{Q}$$

I. Rappels sur que l'on appelle champ moyen

Champ moyen :

- \mathcal{Q} est la famille des distributions factorisables sous la forme $\prod_{i=1}^n Q_i(x_i)$
- la solution de min KL consiste à fixer l'influence des voisins à sa valeur moyenne
- En général une équation de point fixe à résoudre.

I. Rappels sur ce que l'on appelle champ moyen

Soit

$$P(x_1, \dots, x_n) = \frac{1}{Z} \exp\{-H(x_1, \dots, x_n)\}$$

alors

$$Q_i^{CM}(x_i) \propto \exp\{E_Q[-H(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n)]\}$$

et si H s'écrit comme une somme de potentiels linéaires (ex modèle de Potts)

$$Q_i^{CM}(x_i) = P(x_i | X_{N(i)}) = E^{CM}[x_{N(i)}]$$

→ les marginales ne sont pas conservées

I. Rappels sur que l'on appelle champ moyen

Quelques remarques:

- l'étape de min KL n'a pas toujours une solution analytique si on choisit Q plus compliquée (Belief Propagation)
- que retrouve-on **a priori** dans les 3 applications étudiées?
 - estimation/inférence pour modèles graphiques (Florence) : étapes 1 et 2
 - optimisation d'un PDMG (Régis) : 1 et 2 mais sans équation de point fixe
 - recherche de l'équilibre d'un processus de contact (Alain) : étape 1 uniquement

II. CM comme borne inf de la vraisemblance

$x = (x_1, x_2, \dots, x_n)$: variables cachées, y observations, θ : paramètre(s) du modèle.

$$\begin{aligned}\ln P(y | \theta) &= \ln \sum_x P(y, x | \theta) \\ &= \ln \sum_x Q(x) \frac{P(y, x | \theta)}{Q(x)} \\ &= \ln E_Q \left[\frac{P(y, x | \theta)}{Q(x)} \right] \\ &\geq E_Q \left[\ln \frac{P(y, x | \theta)}{Q(x)} \right] \\ &= \ln P(y | \theta) - KL(Q(x) | P(x | y, \theta))\end{aligned}$$

II. CM comme borne inf de la vraisemblance

Lien entre précision sur la vraisemblance et précision sur les marginales?

Illustration sur un petit modèle (cf Advance Mean Field Methods, chap 10)

- $n = 2, x_i \in \{0, 1\}$
- pas de données observées
- la log vraisemblance est nulle : $\ln \sum_x P(y, x | \theta) = \ln \sum_x P(x | \theta) = \ln(1)$
- $P(0, 0) = (1 - p)/2; P(0, 1) = p/2; P(1, 0) = p/2; P(1, 1) = (1 - p)/2$
- les marginales $P(0) = P(1) = 1/2$
- si $p = 1/2$, loi de $(x_1, x_2) =$ loi uniforme

Lien entre précision sur la vraisemblance et précision sur les marginales?

$$\ln P(y | \theta) = \ln P(y | \theta) - KL\left(\prod_i Q(x_i) \mid P(x | y, \theta)\right)$$

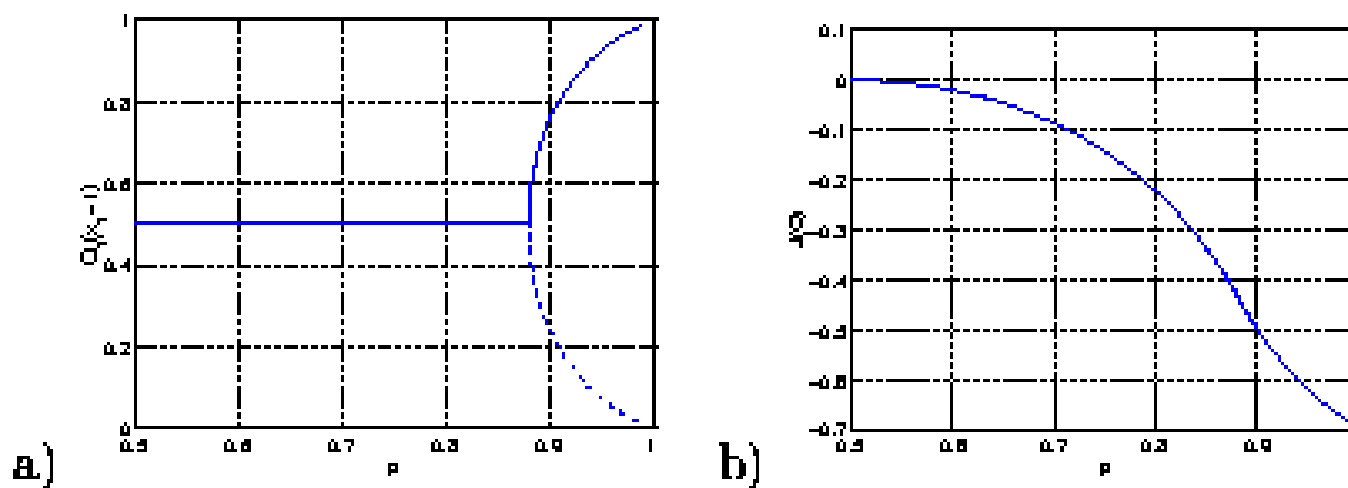


Figure 2: a) $Q_1(x_1 = 1)$ resulting from symmetry breaking as a function of the parameter p . The dashed line represents the alternative solution resulting from different initialization. b) the lower bound $J(Q)$ as a function of p .

III. CM comme borne inf de la constante de normalisation d'un champ de Gibbs

- écrivons la distribution de (x_1, \dots, x_n) sous la forme :

$$P(x_1, \dots, x_n) = \frac{1}{Z} \exp\{-H(x_1, \dots, x_n)\}$$

- et une distribution Q dans la famille \mathcal{Q} des distributions factorisables de la forme

$$\begin{aligned} Q(x_1, \dots, x_n) &= \frac{1}{Z^Q} \exp\{-H^Q(x_1, \dots, x_n)\} \\ &= \prod_{i=1}^n \frac{1}{Z_i^Q} \exp\{-H^Q(x_i)\} \end{aligned}$$

- soit $\Delta H = H - H^Q$,

$$\begin{aligned}
 Z &= \sum_x \exp\{-H(x)\} \\
 &= \sum_x \exp\{-H^Q(x) - \Delta H(x)\} \\
 &= Z^Q \sum_x \frac{\exp\{-H^Q(x)\}}{Z^Q} \exp\{-\Delta H(x)\} \\
 &= Z^Q E_Q[\exp\{-\Delta H(x)\}] \\
 &\geq Z^Q \exp\{E_Q[-\Delta H(x)]\} \\
 &= Z^Q \exp\{E_Q[-H(x) + H^Q(x)]\}
 \end{aligned}$$

- Maximiser le terme de droite pour Q dans \mathcal{Q} revient à minimiser $KL(Q | P)$

IV. CM et champ de Markov : un encadrement grossier de $KL(Q | P)$

Xing E., Jordan M. and Russel S., Graph partition strategies for generalized mean field inference, Uncertainty in Artificial Intelligence, 2004.

Supposons une grille régulière avec 4 plus proches voisins et

$$P(x_1, \dots, x_n) = \frac{1}{Z} \exp\left\{ \sum_i \theta_i H_i(x_i) + \sum_{i \sim j} \theta_{ij} H_{ij}(x_{ij}) \right\}$$

alors

$$aW \leq KL(Q | P) \leq bW$$

avec

$$- W = \sum_i \theta_i + \sum_{i \sim j} \theta_{ij}$$

- a et b coefficients fonctions des potentiels max et min

IV. CM et champ de Markov : un encadrement grossier de $KL(Q | P)$

Cas du modèle de Potts

$$P(x_1, \dots, x_n) = \frac{1}{Z} \exp\left\{ \sum_{i \sim j} \theta \mathbb{1}_{x_i = x_j} \right\}$$

l'encadrement devient ...

$$0 \leq KL(Q | P) \leq 2 \theta \times \#\text{aretes}$$

Est-il toujours nécessaire d'aller au delà du CM

- Quelques éléments mais question encore ouverte
- CM marche bien lorsque le nombre moyen de voisins est grand, ou la dimension du graphe est grande
- La qualité peut dépendre de l'objectif
- La mise en œuvre des méthodes variationnelles d'ordre supérieur n'est pas toujours possible

Est-il toujours nécessaire d'aller au delà du CM

Quelques exemples :

- estimation de l'équilibre du processus de contact : les méthodes variationnelles d'ordre 2 améliorent grandement pour beaucoup de familles de graphes ... exposé Alain
- **Modèle graphique** : (cf Advance Mean Field Methods, chap 15)
 - CM = Méthode Var d'ordre 1, exact si indépendance ... naïf
 - Bethe = Méthode Var d'ordre 2, exact sur un arbre.
 - graphe quelconque Bethe $>$ MF pour inférence
 - pour Maximum a Posteriori (décision dure) MF peut suffire

Conclusion

- Pas d'encadrement théorique de la KL
- La KL n'est pas forcément le critère intéressant
- En pratique du cas pas cas et de la validation empirique