Un panorama des principales méthodes de calcul de la constante de normalisation d'un champ de Markov

N. Peyrard

Qu'est-ce qu'on veut calculer?

- **Le modèle**
champ de Markov (MRF) : $X = \{X_i, i \in V\}$, $|V| = n$,
$X_i \in \Omega = \{0, \dots K-1\}$
$G = \{V, E\}$, un graphe d'indépendance conditionnelle

$$P(X = x) \propto \prod_c f_c(x_c)$$

f_c, fonction potentielle associée à une clique du graphe

- **L'exemple classique**
Potts 2 couleurs sur une grille

$$\log(f_i(x_i)) = \alpha x_i, \quad \log(f_{ij}(x_i, x_j)) = \beta\delta(x_i, x_j)$$

Qu'est-ce qu'on veut calculer?

- Le graal (un des) : **la constante de normalisation** Z

$$Z = \sum_{x \in \Omega^n} \prod_c f_c(x_c)$$

- **Caratéristiques**
  - problème de type sum-prod
  - nombre de termes exponentiel en le nombre de sites de G

A quoi ça sert de connaître Z?

- **Estimation** des paramètres du modèle : calcul de la vraisemblance
- **Sélection de modèle** : BIC s'écrit comme une ratio de deux constantes de normalisation
- **Inférence** : $P(x_i) = Z_i(x_i)/Z$
- Physique statistique : $\log(Z) =$ minimum de la fonction d'**énergie libre**

Une classification des méthodes

I Méthodes exactes
- I.1 par élimination de variable
- I.2 par parcours d'arbre

II Méthodes approchées
- II.1 Méthode par élimination et décomposition de fonction
- II.2 Méthodes par passage de messages
- II.3 Méthodes par simulation

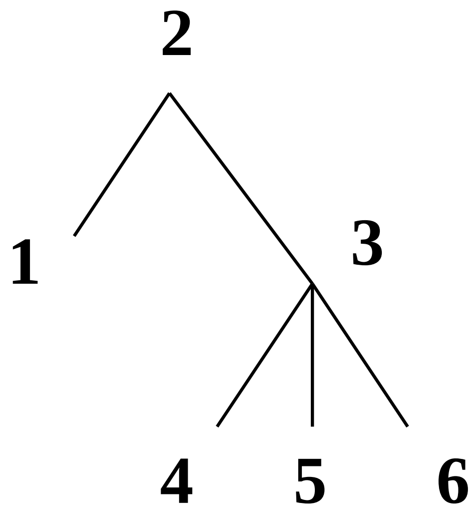
I Méthodes exactes

- Par **élimination de variable**
  - élimination par variable ou par groupe de variables
  - récursivité
  - ordonnancement des variables
  - exploitation de la structure de G
  - programmation dynamique non sérielle
    (cf exposé R. Sabbadin et T. Schiex MSTGA nov. 2010)

- Par **parcours d'arbre**
  - recherche exhaustive
  - séparation (conditionnement) et évaluation

I.1 Principe des méthodes par élimination de variables

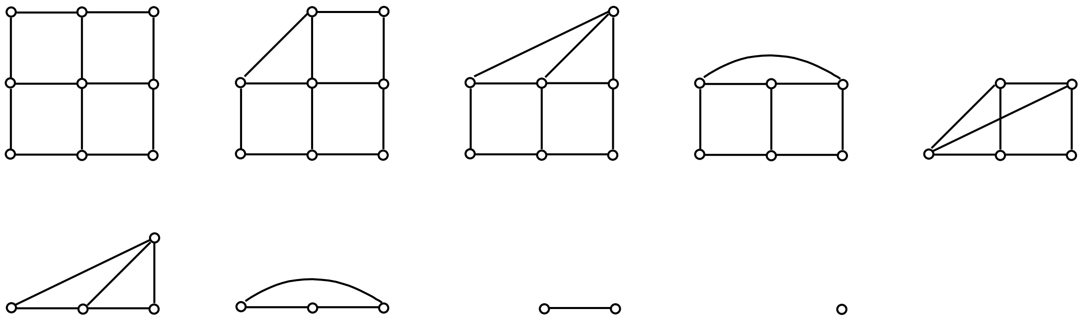
Arbre : tree-width = 1

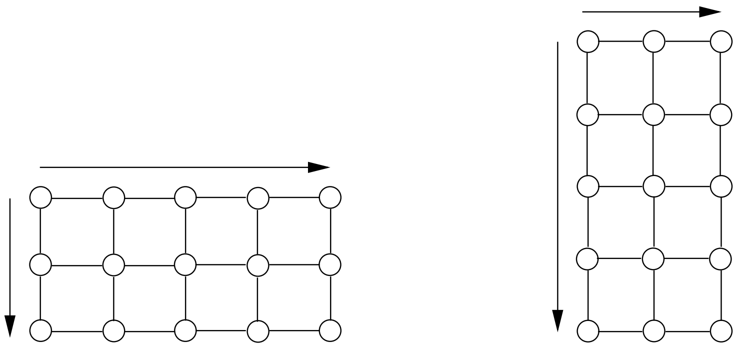


$$Z = \sum_{x_2} f_2 \sum_{x_1} f_1 f_{1,2} \sum_{x_3} f_3 f_{1,3} \sum_{x_4} f_4 f_{3,4} \sum_{x_5} f_5 f_{3,5} \sum_{x_6} f_6 f_{3,6}$$

I.1 Principe des méthodes par élimination de variables

Grille $M \times N$: tree-width = $\min(M, N) - 1$

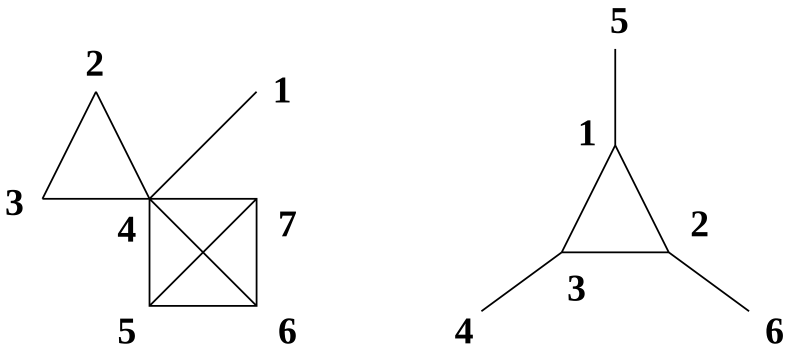




a b

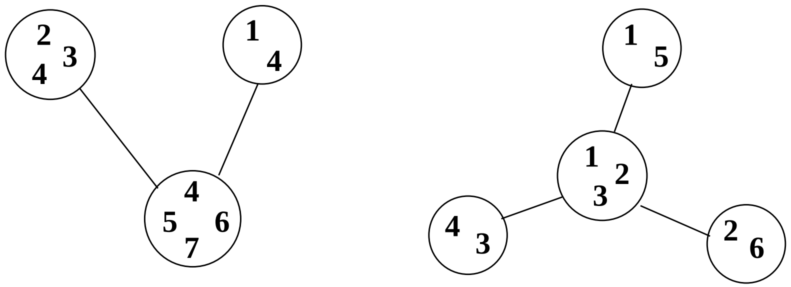
I.1 Principe des méthodes par élimination de variables

Graphe chordal : tree-width = $\max(tailleclique) - 1$



(a) (b)

Arbre de jonction



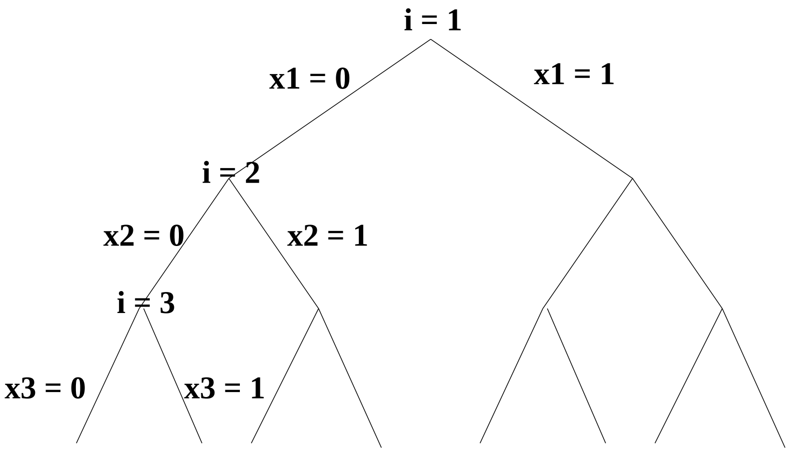
(a) (b)

I.1 Quelques références

- **Historiques** : élimination de variables sur graphe : Bertelè et Brioschi 1972 , Pearl 1988

- **Calcul de Z sur grille** : Pettitt, Friel et Reeves 2003, Reeves et Pettit 2004, Hardouin et Guyon 2010
(cf exposé de C. Hardouin à MSTGA juin 2010)

I. 2 Principe des méthodes par parcours d’arbre



- une feuille contient l’élément de $\sum_{x \in \Omega^n} \prod_c f_c(x_c)$ pour l’instanciation qui mène à la feuille
- $Z =$ somme des valeurs des feuilles
- généralement combiné avec de l’élimination de variable
- aime bien le déterminisme

I.1 Quelques références

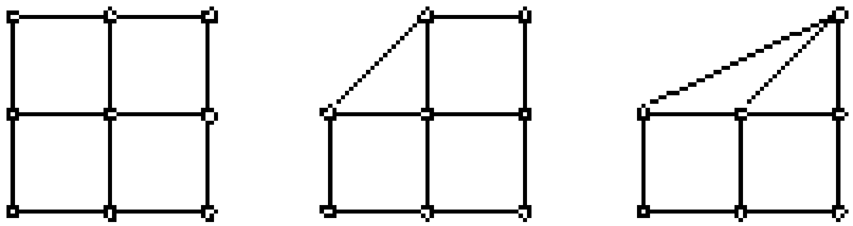
- **Historiques** : Pearl 1984

- **Inférence dans les modèles graphiques**
  - Dechter 1999 : élimination de variable et conditionnement
  - Dechter et Mateescu 2007 : arbre de jonction et parcours d'arbre
  - Gogate et Domingos 2010 : conditionnement sur des groupes de variables

II Méthodes approchées

II.1 Principe des méthodes par élimination et décomposition de fonction

- Exemple sur la grille



$$Z = \sum_{x_3,\ldots x_9} \prod_{i>2} f_i(x_i) \prod_{i,j>2} f_{i,j}(x_i, x_j) h(x_3, x_4, x_5, x_6)$$

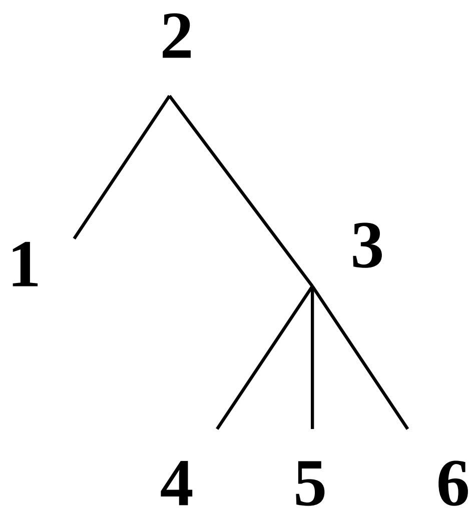
$$\begin{aligned}
h(x_3, x_4, x_5, x_6) &\approx g_{3,4,5}(x_3, x_4, x_5) g_{3,6}(x_3, x_6) \\
\text{ou} &\approx g_{3,4}(x_3, x_4) g_{4,5}(x_5, x_6) \\
\text{ou} &\approx g_{3,4,5}(x_3, x_4, x_5) g_{4,5,6}(x_4, x_5, x_6)
\end{aligned}$$

I.1 Quelques références

- Dechter et Rish 2003 (Mini-Bucket, MB): utilise la forme des fonctions qui ont généré h (en général combiné à d'autres approches)
- Wexler et Meek 2008 (MAS): en théorie minimise des bornes
- Favier et al. 2011 : recherche d'indépendances (exact ou approché)

II.2 Principe des méthodes par passage de message

- une méthode locale, distribuée pour le calcul de toutes les marginales $P(x_i)$ *(cf exposé de F. Forbes à MSTGA sept. 2007)*



$$m_{i\to j}(x_j) = \sum_{x_i} f_{ij}(x_i, x_j) f_i(x_i) \prod_{k\in N(i)\backslash j} m_{k\to i}(x_i)$$

II.2 Principe des méthodes par passage de message

- mais encore ...
  - exact sur arbre (en deux passes)
  - de type élimination de variable
  - approché sur graphe qlq :
    - loopy/iterative BP (LBP): une variable à la fois
    - Generalized BP (GBP): élim par paquets de variables
  - lien avec méthodes variationnelles en Physique Statistique *(cf lecture YFW MSTGA avril 2008)*

II.2 Quelques références

- Yedidia, Freeman et Weiss 2005 : GBP, extension de LBP, les messages sont entre régions
- Wainwright, Jaakkola, Willsky 2005 (Tree-reweighted, TRW): optimise un majorant de Z_Θ parmi une décomposition convexe de Θ selon les spannings tree de G
- Choi et Darwiche 2008 : suppression d'arêtes et apprentissage des marginales par BP, puis extension au rajout d'arêtes

II.3 Principe des méthodes par simulation

- Gibbs Sampling
- Importance sampling
- Utilisation d'une distribution auxiliaire P_0 plus simple et estimation de $\log(\frac{Z}{Z_0})$

II.2 Quelques références

- Potamianos et Goutsias 1997 : 4 méthodes
- Gogate et Domingos 2010 : Importance sampling avec une distribution auxiliaire sur des groupes de variables et non sur les variables

En résumé

En résumé : les plus et les moins

Méthode	Bornes	+	-
exactes	nc	exact	tps/mémoire
messages	TRW : M	bon en pratique	res. théoriques difficiles
élim. + décomp.	MAS : m, M MB : m,M	idem	heuristiques
simulation	non	conv. asymp.	long
naïves	MF : m piecewise : M	très rapide	grossier

En résumé : les boutons à régler

- **élimination + décomposition** : quand décompose-t-on? en combien de sous-fonctions? de quel scope?
- **passage de message** : initialisation? critère d'arrêt? mise à jour séquentielle ou parallèle?
- **simulation** : réglages classiques d'un MCMC. Choix de la distribution auxiliaire pour Importance Sampling

En résumé : quelques impressions

- bibliographie non exhaustive *(cf exposé N. Frial MSTGA nov. 2010)*
- ça commence à percoler entre les disciplines (stat, IA, physique statistique)
- élimination de variables plus développé que parcours d'arbre (surtout LBP, TRW, MB)
- difficile de comparer les méthodes empiriquement : disponibilité code, benchmarks, valeur exacte de Z pas toujours disponible
- difficile de savoir où est-ce qu'il reste de la marge de progression

Des logiciels pour l'inférence dans les modèles graphiques

- libDAI (Joris Mooj): http://people.kyb.tuebingen.mpg.de/jorism/libDAI/
- groupe Automated Reasoning, UCI : http://graphmod.ics.uci.edu/group/Software
- IJGP-SampleSearch (Vibhav Gogate) : http://www.cs.washington.edu/homes/vgogate/ijgp-samplesearch.html
- Toulbar2, SaAb UBIAT : http://mulcyber.toulouse.inra.fr/projects/toulbar2/
- Talya Meltzer : http://www.cs.huji.ac.il/ talyam/inference.html

References

[BB72] U. Bertele and F. Brioschi. *Nonserial Dynamic Programming*. New York academic Press, 1972.

[CD08] Arthur Choi and Adnan Darwiche. Approximating the partition function by deleting and then correcting for model edges. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, pages 79–87, 2008.

[D99] R. Dechter. Bucket elimination: A unifying framework for reasoning. *Artificial Intelligence*, 113:41–85, 1999.

[DM07] R. Dechter and R. Mateescu. And/or search spaces for graphical models. *Artificial Intelligence*, 171:73–106, February 2007.

[DR03] R. Dechter and I. Rish. Mini-buckets: A general scheme for bounded inference. *J. ACM*, 50:107–153, March 2003.

[FdGLS11] A Favier, S de Givry, A Legarra, and T Schiex. Pairwise decomposition for combinatorial optimization in graphical models. 2011.

[GD10] V. Gogate and P. Domingos. Formula-based probabilistic inference. In *26th Conference on Uncertainty in Artificial Intelligence*, 2010.

References

[HG10] Cécile Hardouin and Xavier Guyon. Exact marginals and normalizing constant for Gibbs distributions. *Comptes Rendus de l'Académie des Sciences*, 348(3-4):199–201, 2010.

[P84] J. Pearl. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Reading, MA: Addison-Wesley, 1984.

[P88] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[PFR03] A. N. Pettitt, N. Friel, and R. Reeves. Efficient calculation of the normalizing constant of the autologistic and related models on the cylinder and lattice. *Royal Statistical Society Journal: Series B (Statistical Methodology)*, 65(1):235–246, 2003.

[PG97] G. Potamianos and J. Goutsias. Stochastic approximation algorithms for partition function estimation of Gibbs random fields. *IEEE Transactions on Information Theory*, 43(6):1948–1965, 1997.

[RP04] R. Reeves and A. N. Pettitt. Efficient recursions for general factorisable models. *Biometrika*, 91(3):751–757, 2004.

References

[WJW05] M J Wainwright, Tommi S Jaakkola, and Alan S Willsky. A new class of upper bounds on the log partition function. *IEEE Transactions on Information Theory*, 51(7):2313–2335, 2005.

[WM08] Ydo Wexler and Christopher Meek. Mas: a multiplicative approximation scheme for probabilistic inference. In *NIPS*, pages 1761–1768, Vancouver, Canada, 2008.

[YFW05] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *Information Theory, IEEE Transactions on*, 51(7):2282–2312, 2005.