SUR L'ANALOGIE ENTRE LA MECANIQUE STATISTIQUE ET LES SYSTEMES EXPERTS

Par

Hung T.Nguyen

Department of Mathematical Sciences

New Mexico State University

Las Cruces, New Mexico 88003-USA.

0. Dans cette Note, nous allons mettre en évidence l'analogie entre l'analyse de la mécanique statistique et celle des systèmes experts, et cela en vue d'étudier les représentations des connaissances et les modes d'inférence en systèmes experts.
Chemin faisant, nous allons mettre en évidence aussi le role naturel que joue le concept d'un ensemble aléatoire en engendrant des mesures d'incertitude non-additives, e.g.fonctions de croyance, mesures de possibilité.

1. Des travaux récents ont montré, d'une façon assez éloquente, qu'il est très utile d'utiliser les idées et techniques de la mécanique statistique pour étudier les systèmes experts (e.g. Geman et Graffigne, 1986, Pearl, 1986, Kirkpatrick et al., 1983, Lauritzen et Spiegelhalter, 1988). Ceux-ci

ont notamment mis l'accent sur la représentation des connaissances (structures graphiques), et sur l'inférence probabiliste en exploitant les représentations locales des structures Markoviennes. Aussi, dans ce context, une nouvelle méthode d'optimization, celle du recuit simulé, a été proposée. D'autre part, ces travaux peuvent etre interpretés comme efforts pour défendre l'usage des probabilités (Bayésiennes) et leurs calculs rigoureux dans le domaine de l'Intelligence Artificielle.A cet effet, voir aussi l'article de Lindley (1982) et Goodman et Nguyen (1985,p 558-567) !

Ici il ne s'agit pas de prendre une position, mais bien au contraire, notre but vise à :

(i) Etendre l'analogie entre la mécanique statistique et l'inférence probabiliste aux autres mesures d'incertitude,

(ii) Soulever des problèmes analogues pour autres méthodes d'inférence, e.g. représentations locales,calculations rapides sur les graphes,

(iii) Compléter l'analyse probabiliste dans le cas ou l'imprécision est dominante.

Ceci étant, nous partons de l'analyse probabiliste.

2. Commencons brivièvement par dégager la notion d'un système expert ! En absence d'une définition mathématique, contentons de "décrire" un tel système comme suit. Une machine pour l'aide a la décision est a construire.Comme cette machine sera appelée "intelligente", il doit s'agir d'un problème de décision devant l'incertitude.En outre, elle doit se comporter comme un expert ! Deux composantes principales sont donc :

a) Les données ,

b) Les méthodes d'inférence .

La situation est plus compliquée par rapport à la théorie des décisions statistiques. En effet, tout d'abord, les données sont fournies par les experts (qualitatives et quantitatives) et nécessitent une certaine façon de les représenter.Les méthodes d'inférence reposent sur le calcul de la mesure d'incertitude choisie. Ainsi, si les probabilités Bayésiennes sont utilisées pour modeler l'incertitude dans les opinions des experts, on est conduit à l'inférence probabiliste.En pratique, le choix d'une mesure d'incertitude peut etre guidé par des considérations d'implémentation, e.g. données manquées, imprécision, calculs rapides, ect...

De façon précise, soit $\{ X_v ,v \in V \}$ une collection de variables d'un domaine de connaissance, disons médicale. L'ensemble V est fini mais pas nécesairement un sous ensemble de la droite réelle. De sorte que cette collection de variables est un champ aléatoire en général. Les experts vont fournir non seulement des informations quantitatives sur ces variables, e.g. densités conditionelles, mais aussi des relations entre ces variables.Donc l'ensemble des indices V va posséder une certaine structure plus précise.Prenons le cas le plus important, à savoir V est l'ensemble des sommets d'un graphe (orienté) G = (V,E) où E est l'ensemble des arcs.Il s'agit de construire la probabilité conjointe des X_v à partir des données d'une telle façon que la propagation de l 'incertitude dans ce réseau peut se faire localement, c'est-à-dire rapidement .D'autre part, il arrive souvent que chaque variable X_v n'est influencée que par des variables "voisines". Autrement dit on peut s'apercevoir une certaine sorte d'indépendance conditionelle entre les variables. C'est là l'analogie avec la mécanique statistique (voir e.g. Preston, 1974, 1976, Rozanov, 1982)

3. Considérons le cas des variables binaires (oui,non ou 0-1). L'espace des configurations est alors identifié à l'ensemble de toutes les parties de V, P(V).Le champ aléatoire $\{X_v,v \in V\}$ est gouverné par une probabilité sur P(V). Mais comme P(V) est fini, une telle probabilité est identifiée à sa

densité μ qui est une fonction définie sur P(V) telle que $\Sigma_{A \subset V}\, \mu(A) = 1$.μ est précisément la densité d'un ensemble aléatoire à valeurs dans P(V). On voit déjà qu'une fonction de croyance sur V peut etre définie par $Bel(B) = \Sigma_{A \subset B}\, \mu(A)$ (pour un exposé excéllent de la théorie des fonctions de croyance, voir Wasserman, 1987).

Par analogie avec la mécanique statistique, on peut considérer une fonction φ , positive,définie sur P(V) telle que $\varphi(\phi) = 0$, appelée énergie, et prendre:

$$\mu(A) = (1/Z)\exp(\varphi(A))$$

où Z est la fonction de partition $Z = \Sigma_{B \subset V} \exp(\varphi(B))$, notons que la densité μ ainsi définie est strictement positive, $\mu(\phi) > 0$.

En mécanique statistique, l'énergie φ est la somme des potentiels dus aux sous systèmes,notés J_φ ($J_\varphi(B)$ représente le potentiel du aux éléments de B) de sorte que $\varphi(A) = \Sigma_{B \subset A} J_\varphi(B)$, et par la formule d'inversion de Mobius, on a $J_\varphi(B) = \Sigma_{C \subset B} (-1)^{|B-C|}\varphi(C)$, où $|A|$ est la cardinalité de A. Quand le graphe a seulement des interactions voisines, la probabilité conjointe est Markovienne, et celà se traduit en termes des graphes en prenant un potentiel(de Gibbs) non nul seulement sur les sous ensembles de V qui sont complets,i.e., deux sommets quelconque d'un tel sous ensemble sont "voisins".Dans un tel cas, on a une représentation locale de μ à partir des densités conditionelles sur les arcs du graphe (Lauritzen et Spiegelhalter, 1988). Pour l'inférence, on a besoin des transformations d'une représentation locale à une autre préservant la propriété de Markov. Notons que comme chaque variable aléatoire X engendre une tribu $\sigma(X)$, la propriété de Markov peut s'exprimer d'une manière générale en termes d'une collection des tribus de la forme σ_S ,$S \subset V$. La propriété de Markov est préservée par temps d'arret. Or ici le "temps" devra etre un ensemble aléatoire satisfaisant à une certaine condition de mesurabilité (en temps que fonction multivoque, voir e.g. Rozanov, 1982.).Autrement dit, la propriété forte de Markov s'exprime en termes des ensembles aléatoires. .

4. Revenons aux fonctions de croyance. Il est bien connu que cette théorie peut etre interprétée comme distributions des ensembles aléatoires (e.g. Goodman et Nguyen, 1985). L'article récent de Hummel et Landy (1988) ne contient pas des éléments nouveaux à cet effet, en remarquant que,d'une part, un espace des opinions des experts n'est simplement qu'un espace pondéré des observateurs, comme en théorie de l'information généralisée de Kampé de Fériet,et d'autre part, les opinions Booléennes forment un ensemble aléatoire ! . Donnons nous maintenant une autre interprétation de cette théorie en langage de la mécanique statistique.

Soit V un ensemble fini. Soit m une fonction sur $\mathbf{P}(V)$ telle que $\Sigma_{A \subset V}$ m(A) = 1. Considérons Bel(A) = $\Sigma_{B \subset A}$ m(B) . Imaginons qu'a chaque point v de V, une particule est présente ou absente (modèle d'Ising). Interprétons m(.) comme potentiels d'interaction, de sorte que Bel(.) est une énergie du système. La probabilité conjointe ayant pour énergie Bel(.) est donc :

$\mu(A) = (1/Z)\exp(\mathrm{Bel}(A))$! (Notons que μ est la densité de cette probabilité sur $\mathbf{P}(V)$).Si de plus,m(A) est non nul seulement pour A complet, il est alors bien connu que μ est Markov (voir,e.g. Preston,1974,p 5). De cette façon, on peut donc revenir à l'analyse probabiliste comme dans Lauritzen et Spiegelhalter (1988). Notons que la propriété de Markov de μ est nécessaire pour les calculs locaux quand on passe d'une représentation locale a une autre pour l'inférence.Voir aussi Schenoy et Shafer(1986).

La théorie des fonctions de croyance est basée essentiellement sur l'argument de type suivant : étant donné une "évidence" e, le degré de croyance à un évènement A est Bel(A|e) qui peut etre interprété formellement

comme "la probabilité que l'évidence e implique A ", en symbole P(e → A). Supposons que P est une probabilité sur V (avec tribu $\mathbf{P}$(V)) et que l'évidence e est un sous ensemble de V. Aussi supposons que l'implication → est définie par e→ A = e' or A. Alors il est facile de voir que :
P(A|e) est différent de P(e→ A) en général.Pour une interprétation formelle, il est peut etre utile d'étudier des implications logiques appropriées. D'autre part, on construit souvent des fonctions de croyance simples basées sur l'évidence du type : soit B un évènement avec P(B) = α . On prend m(B) = α ,m(V) = 1-α . En pratique, l'information est conditionelle, disons de la forme P(A|B) = α .Alors comment construit-t-on une fonction de croyance associée ? Cette difficulté est due principalement au fait que l'évènement conditionel (A|B) n'est pas rigoureusement défini en théorie des probabilités. Rappelons que dans les fondements de la théorie mathématique des probabilités donnés par Komogorov, la notion d'évènement conditionel n'est défini que par une considération numérique,i.e. en termes des probabilités conditionelles. Pour l'histoire de ce problème ainsi que pour une étude nouvelle du conditionellement, voir le travail récent de Goodman et Nguyen (1987) dans lequel l'évènement conditionel (A|B), pour A,B éléments d'une algèbre de Boole, est défini comme une classe d'équivalence d'évènements, et celà sans faire appel à aucune mesure sur l'algèbre..

5. De ceux qui précèdent, on voit que la mécanique statistique a fourni des idées fécondes à l'analyse des systèmes experts en ce qui concerne les probabilités Bayésiennes, à savoir structures graphiques, champs aléatoires de Markov, temps d'arret, représentations locales (en vue de la propagation rapide de l'incertitude à travers le réseau), etc... Meme dans ce cas, comme ont signalé Lauritzen et Spiegelhalter (1988), les probabilités (conditionelles) sur les arcs du graphe en question peuvent etre

imprécises,on pourra penser aux mesures de possibilité de Zadeh pour modeler ce genre d'incertitude.

A un niveau plus général, étant donné la complexité d'un problème réel, l'imprécision peut apparaitre dans la représentation de la connaissance aussi bien que dans les données quantitatives. Une étude englobant les graphes flous et probabilités floues est alors désirable. Les interactions voisines des variables doivent se traduire d'une manière indépendante de la mesure choisie. Peut -t-on développer des représentations locales semblables au cas Bayésien ?

Il est bien connu qu'il y a aussi une connection entre les mesures de possibilité de Zadeh et les ensembles aléatoires (e.g. Goodman et Nguyen, 1985). D'autre part, la règle de combinaison de Demspter-Shafer est utilisée seulement quand les fonctions de croyance sont indépendantes, i.e. les ensembles aléatoires associés sont indépendants, on pourra s'appuiser sur cette notion commune qui est l'ensemble aléatoire pour faire une étude générale. Dans cette direction, quand it s'agit du problème des données manquées, on pourra faire appel aux résultats assez récents de la statistique des ensembles aléatoires ou quand il s'agit des méthodes d'estimation basées sur l'entropie du système, à la nouvelle technique d'optimization appelée recuit simulé, Kirkpatrick et al., 1983..

6.Bibliographie.

[1] Berge, C.(1970) **Graphes et Hypergraphes**. Dunod, Paris.

[2] Geman, S. et Graffigne, C.(1987) Markov random field image models and their applications to computer vision. in **Proceedings of the Inter.Congress Math.1986** (A.M.Gleason,Ed.), Amer.Math.Soc., Providence.

[3] Goodman, I.R. et Nguyen, H.T.(1985) **Uncertainty Models for Knowledge-Based Systems**. North-Holland, Amsterdam.

[4] Goodman, I.R. et Nguyen, H.T. (1987) **A Theory of Measure-Free Conditioning**. Preprint.

[5] Hummel, R.A. et Landy,M.S.(1988) A Statistical Viewpoint on the Theory of Evidence. **IEEE Trans. Pattern Anal. and Machine Intell.**, vol 10, no 2, 235-247.

[6] Kirkpatrick, S., Gelatt, C.D. et Vecchi,M.P. (1983) Optimization by Simulated Annealing. **Science**, vol 220 ,671-680.

[7] Lauritzen, S.L. et Spiegelhalter, D.J. (1988) Local computations with probabilities on graphical structures and their application to expert systems. To appear in **J.R.Statist.Soc**, Series B.

[8] Lindley, D.V. (1982) Scoring rules and the inevitability of probability. **Inter.Statist.Rev.**, 50, 1-26.

[9] Pearl, J (1986) Fusion, propagation and structuring in belief networks. **Artificial Intelligence**, 29, 241-288.

[10] Preston, C.J. (1974) **Gibbs States on Countable Sets**. Cambridge Univ. Press, London.

[11] Preston, C.J. (1976) **Random Fields**. Lecture Notes in Math., no 534, Springer-Verlag, New York.

[12] Rozonov, Y.A. (1982) **Markov Random Fields**. Springer-Verlag, New York.

[13] Schenoy, P.P. et Shafer, G. (1986) Propagating belief functions with local computations. **IEEE Expert**, 43-51.

[14] Wasserman, L.A. (1987) **Some Applications of Belief Functions to Statistical Inference**. Ph.D. Thesis, Univ. of Toronto, Canada.