

Examen du module MGDE

Durée : 3 heures

Exercice 1

Soit 5 variables aléatoires X_1, X_2, X_3, X_4 et X_5 . On suppose qu'elles vérifient les relations d'indépendance conditionnelles suivantes :

$$X_3 \perp\!\!\!\perp X_1 \mid X_2; X_4 \perp\!\!\!\perp (X_1, X_2) \mid X_3; X_5 \perp\!\!\!\perp (X_1, X_2, X_3) \mid X_4.$$

Q 1.1 Montrer successivement que :

$$\begin{aligned} &X_4 \perp\!\!\!\perp X_1 \mid (X_2, X_3); X_1 \perp\!\!\!\perp (X_3, X_4) \mid X_2; (X_3, X_4, X_5) \perp\!\!\!\perp X_1 \mid X_2; \\ &(X_1, X_2) \perp\!\!\!\perp (X_4, X_5) \mid X_3; X_1 \perp\!\!\!\perp X_5 \mid (X_2, X_3, X_4); X_1 \perp\!\!\!\perp X_5 \mid (X_2, X_3). \end{aligned}$$

Q 1.2 Tracer le graphe d'indépendance obtenu à partir de l'ordre d'énumération naturel des v.a. et montrer que l'on peut retrouver les résultats du a) à partir du critère de d-séparation.

Exercice 2

Soit \succsim une relation de préférence sur $A \times B \times C \times D \times E \times F \times G \times H \times I \times J$ représentable par la fonction d'utilité :

$$u_1(A, B) + u_2(E, G) + u_3(B, C) + u_4(A, C) + u_5(C, D, E, F) + u_6(G, H) + u_7(G, I) + u_8(B, D, J).$$

Q 2.1 Construisez l'arbre GAI obtenu grâce à la séquence d'élimination $A, H, I, G, J, E, F, B, C$. Vous indiquerez à côté de chacune des cliques les utilités stockées dans celles-ci.

Q 2.2 Supposons que le couple (B, D) soit instancié à (b^0, d^0) . Quelles simplifications peut-on apporter au réseau GAI à partir de cette information ? Vous dessinerez le réseau simplifié et, là encore, vous indiquerez à côté de chacune des cliques les utilités stockées dans celles-ci.

Q 2.3 On recherche l'élément X préféré sur l'hyperplan $(B, D) = (b^0, d^0)$. Indiquez sur le réseau GAI de la question 2 le contenu des messages que vous faites transiter pour obtenir X .

Q 2.4 On rajoute à $(B, D) = (b^0, d^0)$ la contrainte supplémentaire $J = j^0$. Quelles modifications doit-on apporter aux messages de la question précédente pour obtenir l'élément préféré vérifiant ces contraintes ?

Exercice 3

On cherche à prévoir l'impact sur les salaires des informaticiens d'une région d'une amélioration de la qualité de l'enseignement. Les hypothèses sont les suivantes :

- Le niveau de salaire, variable S , dépend directement (dépendance causale) du niveau du diplôme, variable D , et de la demande du marché du travail du secteur, variable M ;
- la qualité de l'enseignement, variable E , a une influence causale sur D , c-à-d sur le niveau de diplôme atteint ;

- le contexte régional, variable U , non observable parce que trop difficile à caractériser, influe causalement à la fois sur les variables E et M ;
- toutes les autres variables sont observées ; on connaît donc les probabilités $p(m, e, d, s)$ mais pas $p(u, m, e, d, s)$.

Q 3.1 Représenter le graphe causal G_0 de sommets U, M, E, D et S exprimant les hypothèses précédentes ; quelle est la décomposition correspondante de $p(u, m, e, d, s)$?

Q 3.2 Représenter le graphe causal G_1 résultant de G_0 par une intervention $E \leftarrow e_0$. Expliquer pourquoi la loi de (U, M, D, S) conditionnelle à cette intervention a pour expression

$$p(u, m, d, s \| e_0) = p(u) \cdot p(m/u) \cdot p(d/e_0) \cdot p(s/m, d);$$

en déduire que

$$p(s \| e_0) = \sum_d p(d/e_0) \cdot \sum_u p(u) \cdot \sum_m p(m/u) \cdot p(s/m, d).$$

Q 3.3 Montrer que les conditions de validité de la formule du blocage aval sont réunies pour permettre l'identification de l'effet causal de E sur S .

Q 3.4 Utilisant la relation

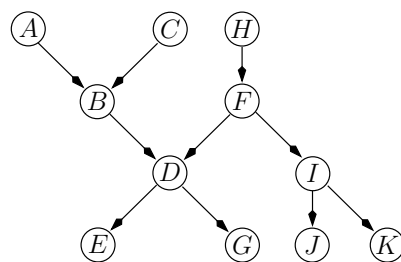
$$p(u) = \sum_e p(u/e) \cdot p(e),$$

ainsi que les relations $p(u/d, e) = p(u/e)$, $p(m/u, d, e) = p(m/u)$ et $p(s/u, m, d, e) = p(s/m, d)$, que l'on justifiera, montrer que

$$p(s||e_0) = \sum_d p(d/e_0) \cdot \sum_e p(s/d, e) \cdot p(e).$$

Exercice 4

Soit le réseau bayésien ci-dessous :



Q 4.1 Moralisez ce réseau.

Q 4.2 Triangulez le graphe moral en utilisant la séquence d'élimination suivante : A, K, I, J, G, F . Vous indiquerez pour chaque nœud éliminé le graphe non orienté obtenu après élimination.

Q 4.3 Dessinez un arbre de jonction correspondant à cette séquence d'élimination et indiquez à côté des cliques les probabilités conditionnelles que vous stockerez dans ces cliques.

Q 4.4 Indiquez les contenus des messages transitant dans les deux sens des arêtes sur chaque séparateur pour le calcul des probabilités *a priori* par l'algorithme de Shafer-Shenoy.

Q 4.5 Dans un arbre de jonction, pour tout couple de cliques voisines C_i et C_j d'intersection S_{ij} , les variables de $C_i \setminus S_{ij}$ sont indépendantes de $C_j \setminus S_{ij}$ conditionnellement à S_{ij} . S_{ij} sépare l'arbre de jonction en deux sous-arbres T_1 et T_2 . Montrez que les variables de $T_1 \setminus S_{ij}$ sont indépendantes de $T_2 \setminus S_{ij}$ conditionnellement à S_{ij} .

Q 4.6 En utilisant la d -séparation, montrez que la propagation d'une information e_A concernant A dans l'arbre de jonction obtenu dans la question 4 ne nécessite le calcul que de deux nouveaux messages.