

**IFT3335 - Exercice 2**  
**À rendre au plus tard le 29 mars avant 23 :59**

Cet exercice compte pour 10% de la note globale du cours. C'est à remettre sur Studium seul ou en groupe de deux personnes. Une pénalité de 20% sera appliquée pour chaque jour de retard.

**Question 1. (15%)**

Formuler les phrases suivantes en logique du premier ordre:

P1. Tout le monde est aimé par quelqu'un.

P2. Il y a des gens qui aiment tout le monde.

P3. Tout le monde aime quelqu'un.

Essayez de démontrer, en utilisant la preuve par réfutation, en appliquant la règle de résolution, si chacune de ces phrases entraîne une autre phrase. C'est-à-dire de tenter de prouver si  $P1 \vdash P2$ ,  $P1 \vdash P3$ ,  $P2 \vdash P1$ ,  $P2 \vdash P3$ ,  $P3 \vdash P1$ ,  $P3 \vdash P2$ .

**Question 2. (10%)**

Est-ce que le raisonnement suivant est correct?

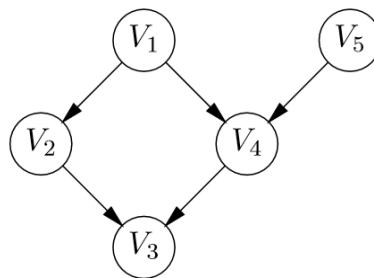
Chaque chose a une cause. Donc, il y a une cause pour toutes les choses.

Démontrer formellement en logique votre réponse. Pour cela, traduisez ces phrases en logique et utilisez une méthode de preuve pour le prouver.

Note : Il y a une ambiguïté (voulue) dans la deuxième phrase. Essayez avec les différentes interprétations de cette phrase.

**Question 3. (15%)**

Cette question concerne la notion d'indépendance (conditionnelle). Supposons le réseau Bayésien suivant :



Est-ce que chacune des équations suivantes est vraie ou fausse ? Expliquer brièvement pourquoi (référez-vous à la sémantique locale du réseau Bayésien et à la notion de couverture Markov).

1.  $P(V3 | V5) = P(V3)$
2.  $P(V5) = P(V5 | V1)$
3.  $P(V5 | V4, V1) = P(V5 | V4)$
4.  $P(V3 | V2, V4, V1) = P(V3 | V2, V4)$
5.  $P(V3 | V2, V1) = P(V3 | V2)$
6.  $P(V3 | V4, V5) = P(V3 | V4)$
7.  $P(V2 | V1, V3, V4, V5) = P(V2 | V1, V3, V4)$
8.  $P(V2 | V1, V4) = P(V2 | V1)$
9.  $P(V4 | V1, V5, V2) = P(V4 | V1, V5)$
10.  $P(V4 | V1, V5, V3) = P(V4 | V1, V5)$

#### Question 4. (15%)

Considérer le problème suivant (tiré du livre de Brachman et Levesque) :

Metastatic cancer is a possible cause of a brain tumor and is also an explanation for an increased total serum calcium. In turn, either of these could cause a patient to fall into occasional coma. Severe headache could also be explained by a brain tumor.

Représenter ces liens dans un réseau Bayésien, en utilisant les variables suivantes :

M – Metastatic cancer

B – Brain tumor

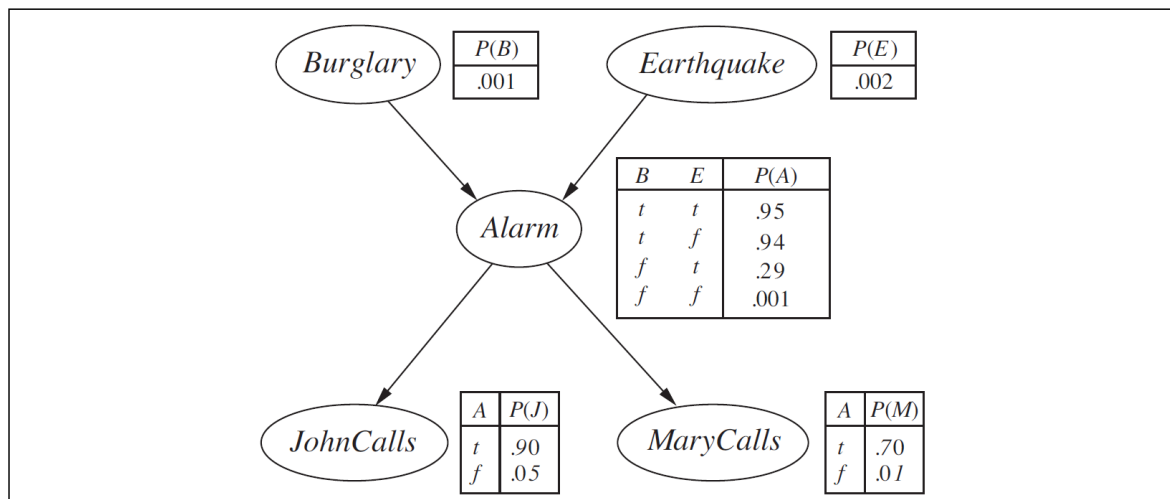
S – Increased total serum calcium

C – Occasional coma

H – Severe headache

#### Question 5. (25%)

Considérer le réseau qu'on a vu en cours, qui est reproduit dans la figure suivante :



a). Calculer les valeurs exactes de  $P(B|j, \neg e)$  (c'est-à-dire les probabilités d'avoir ou non le vol étant donné que John a appelé et qu'il n'y a pas de tremblement de terre) avec la méthode par énumération. Pensez à éliminer des variables non-pertinentes.

b). Implanter en Python ou en Java la méthode d'échantillonnage avec rejet et la méthode d'échantillonnage avec pondération pour estimer les probabilités de la question a). Votre implantation n'a pas besoin d'être générale. Elle doit seulement fonctionner pour ce réseau et pour ces probabilités. Vous devez utiliser une fonction de génération de nombre aléatoire selon la probabilité pour générer une valeur d'une variable.

Lancer ces programmes pour tirer respectivement 100, 1000 et 10 000 échantillons, et calculer  $P(B|j, \neg e)$  selon les échantillons obtenus. Comparer ces estimations avec la réponse exacte que vous avez obtenue dans la question a).

(Note : vous devez rendre ces programmes aussi dans votre remise.)

### Question 6. (20%)

Sur un site d'achat d'un produit sur l'internet, on a collecté les cas suivants qui reflètent les achats ou non des utilisateurs et les caractéristiques (attributs) de ces utilisateurs:

ID	Sexe	Âge	État civil	Revenu	Achat
1	H	18-35	Marié	Moyen	Non
2	H	<18	Célibataire	Faible	Non
3	H	>35	Marié	Elevé	Oui
4	F	<18	Célibataire	Moyen	Non
5	H	18-35	Célibataire	Moyen	Non
6	F	18-35	Célibataire	Elevé	Oui
7	F	18-35	Marié	Faible	Non
8	H	18-35	Marié	Elevé	Oui
9	H	>35	Célibataire	Faible	Oui
10	F	<18	Célibataire	Moyen	Non
11	F	>35	Célibataire	Moyen	Oui
12	F	>35	Marié	Elevé	Oui
13	H	18-35	Célibataire	Faible	Non
14	F	18-35	Marié	Moyen	Oui

1. On voudrait construire un arbre de décision pour déterminer si un client va acheter le produit ou non selon ses caractéristiques. Pour ce faire, on utilise le gain d'information de chaque attribut. Construisez cet arbre de décision, et montrez vos calculs de gains d'information.  
Classez le nouveau cas <F, 18-35, Célibataire, Moyen> selon cet arbre. Est-ce que le client va acheter ou non le produit?  
Si vous observez des problèmes dans l'application de l'arbre pour ce cas, discutez sur ces problèmes.
2. Si on utilise Naïve Bayes pour déterminer si un client va acheter, quelles sont les probabilités à déterminer ? Montrez vos calculs pour ces probabilités.  
En utilisant les probabilités que vous obtenez, quelle sera la probabilité d'achat de la cliente ayant les caractéristiques <F, 18-35, Célibataire, Moyen> ?