

## Examen du module de raisonnement dans l'incertain

Durée : 1 heure 30

Documents autorisés : 1 feuille A4 recto-verso

### Exercice 1 (7 points) — Apprentissage

On a appliqué la première phase de l'algorithme PC sur une base de données, qui nous a produit le squelette suivant :

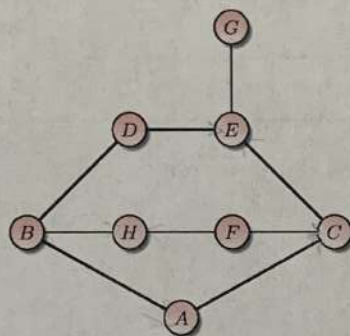


FIGURE 1 – Le squelette appris par PC

Lors de l'apprentissage, PC a déterminé les indépendances suivantes :

$$\begin{array}{llll}
 A \perp\!\!\!\perp D \mid \{B\} & A \perp\!\!\!\perp E \mid \{C, D\} & A \perp\!\!\!\perp F & A \perp\!\!\!\perp G \mid \{C, E\} & A \perp\!\!\!\perp H \mid \{B, F\} \\
 B \perp\!\!\!\perp C \mid \{A\} & B \perp\!\!\!\perp E \mid \{A, D\} & B \perp\!\!\!\perp F \mid \{A, C\} & B \perp\!\!\!\perp G \mid \{A, E\} & \\
 C \perp\!\!\!\perp D \mid \{A, B\} & C \perp\!\!\!\perp G \mid \{E, F\} & C \perp\!\!\!\perp H \mid \{A, F\} & & \\
 D \perp\!\!\!\perp F \mid \{A, C, E\} & D \perp\!\!\!\perp G \mid \{C, E\} & D \perp\!\!\!\perp H \mid \{A, B\} & & \\
 E \perp\!\!\!\perp F \mid \{B, C, G\} & E \perp\!\!\!\perp H \mid \{B, F\} & & & \\
 F \perp\!\!\!\perp G \mid \{E\} & & & & \\
 G \perp\!\!\!\perp H \mid \{B, F\} & & & & 
 \end{array}$$

Q 1.1 Déterminez les v-structures de ce graphe.

Q 1.2 Quel CPDAG obtient-on en appliquant les règles de Meek ?

Q 1.3 Proposez un DAG compatible avec ce CPDAG.

Q 1.4 Sachant que toutes les variables aléatoires sont binaires (booléennes), combien de paramètres doit-on stocker dans les CPT du réseau bayésien correspondant au DAG de la question précédente ?

## Exercice 2 (6.5 points) — Apprentissage bis

Soit trois variables aléatoires booléennes  $A, B, C$  dont on a observé les occurrences ci-dessous :

$A$	$B$	$C$
$a_1$	$b_1$	$c_1$
$a_1$	$b_2$	$c_1$
$a_1$	$b_2$	$c_2$
$a_1$	$b_2$	$c_2$
$a_2$	$b_1$	$c_1$
$a_2$	$b_1$	$c_2$
$a_2$	$b_1$	$c_2$
$a_2$	$b_1$	$c_2$
$a_2$	$b_2$	$c_2$
$a_2$	$b_2$	$c_1$

**Q 2.1** Quels sont les scores log-likelihood en  $A, B, C$  d'un réseau bayésien sans arc et quel est le score du graphe vide (sans arc) ? Pour simplifier, on utilisera la table des logarithmes ci-dessous :

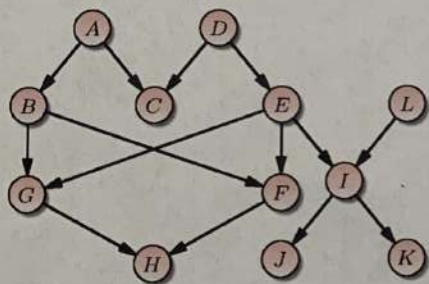
$n$	1	2	3	4	5	6	7	8	9	10
$\log(n)$	0	0.7	1.1	1.4	1.6	1.8	2	2.1	2.2	2.3

**Q 2.2** Quel est le voisinage du graphe vide (sans arc) dans l'algorithme Greedy Hill Climbing ? Quels scores obtient-on pour ce voisinage ? Vous donnerez les formules puis vous les calculerez en utilisant la table des logs fournie plus haut.

**Q 2.3** Quelle modification du graphe Greedy Hill Climbing va-t-il appliquer ? et quel est le score log-likelihood de ce graphe ?

Exercice 3 (4 points) —  $d$ -séparations

On considère le réseau bayésien suivant, de structure  $\mathcal{G}$  :



Qu'est-ce que le critère de  $d$ -séparation permet d'affirmer concernant les propriétés ci-dessous ? Vous justifierez vos réponses.

**Q 3.1**  $\langle A \perp_{\mathcal{G}} L \rangle$  ?

**Q 3.2**  $\langle A \perp_{\mathcal{G}} L | \{J\} \rangle$  ?

**Q 3.3**  $\langle A \perp_{\mathcal{G}} L | \{H, J\} \rangle$  ?

**Q 3.4**  $\langle A \perp_{\mathcal{G}} L | \{B, C, G\} \rangle$  ?

## Exercice 4 (2.5 points) — Ça me scie

Une scierie est implantée sur deux sites, l'un en Savoie et l'autre dans les Pyrénées. Régulièrement, des planches de bois sont sélectionnées au hasard pour effectuer un contrôle qualité. L'entreprise QCat qui effectue ces contrôles fait acheminer les planches jusqu'à son site de Lyon et réalise les mesures nécessaires. Elle en fait acheminer deux fois plus de Savoie que des Pyrénées. Dans le passé, 3/5ème des planches de Savoie étaient de très bonne qualité, 1/5ème étaient de qualité moyenne et 1/5ème de mauvaise qualité. Dans les Pyrénées, 4/5ème étaient de très bonne qualité, 1/5ème étaient de qualité moyenne et aucune de mauvais qualité.

QCat vient d'estimer la qualité d'une planche, qu'elle juge très bonne. Malheureusement, elle a perdu le bordereau d'envoi et ne se rappelle plus si la planche provient de Savoie ou des Pyrénées. Quelle est la probabilité qu'elle vienne de Savoie ? Vous justifierez votre réponse.