Examen du module MGDE

Durée : 3 heures

Exercice 1 (4 points) — *Le dilemne du pauvre amoureux transi*

Théophile est amoureux de Juliette mais pauvre. Il veut lui déclarer sa flamme alors que ses finances lui permettent peu : il doit choisir entre l'emmener au cinéma et lui offrir des fleurs.
Juliette est belle mais simple : si Théophile passe le test et lui offre ce qu'elle attend d'un prince charmant, de manière certaine, elle lui accordera son amour.
Une étude statistique poussée a montré que 50% des femmes belles mais simples veulent un bouquet de fleur comme gage d'amour, 30% préfèrent le cinéma et enfin 20% considèrent que ces 2 cadeaux se valent.

Q 1.1 **(1 point)** Décrire les 3 variables et les relations causales permettant de représenter le dilemne de Théophile. Ne pas oublier les paramètres du modèle.

Remarques : Il n'y a pas de variables cachées. Il manque les paramètres de la loi a priori pour le choix de Théophile.

Les variables :
– $T \in \{B, C\}$: Théophile peut choisir Bouquet ou Cinéma.

- $J \in \{b, c, bc\}$: Juliette peut attendre un $b$ouquet, une place au $c$inema ou les deux (bc).
- $A \in \{a, n\}$: T et J sont des $a$moureux ou ne le sont pas (n).

Le modèle est le graphe G suivant : $T \longrightarrow A \longleftarrow J$
$P(J) = [0.5\ 0.3\ 0.2]$
$P(A = a) = 1$ si ($T = B$ et $J \in \{b, bc\}$) ou si ($T = c$ et $J \in \{c, bc\}$).

Q 1.2 **(1 point)** Théophile a expliqué la situation à son meilleur ami Boris et lui a dit avoir décidé de l'amener au cinéma. Quelque temps plus tard, Boris voit Théophile et Juliette se promener main dans la main et conclut que l'histoire s'est bien terminée. Que peut-il en déduire sur ce que préférait Juliette comme gage d'amour ?

$P(J \mid T = C, A = a) \propto P(A = a \mid J, T = C) \cdot P(J)$ car $P(J \mid T) = P(J)$.
Donc $P(J \mid T = C, A = a) \propto [0\ 0.3\ 0.2]$
$P(J \mid T = C, A = a) = [0\ \frac{3}{5}\ \frac{2}{5}]$

Q 1.3 **(1 point)** Boris ne peut pas s'empêcher de se demander ce qui ce serait passé si Théophile avait fait l'autre choix. Exprimer cette interrogation sous la forme d'une contrefactuelle et résolvez la.

Dans le modèle transformé par $do(C)$, $P_{do(T=C),A=a}(A = a \mid T = B)$ revient à calculer dans G (dont on a supprimé les éventuels parents de T et changer la proba a priori de J par $P(J \mid T = C, A = a)$).
Dans ce réseau, $P_{do(T=C),A=a}(A = a \mid T = B) = P(J = bc) = \frac{2}{5}$. Si Théophile avait fait l'autre choix, il aurait donc eu 60% de rater son idylle.

Q 1.4 **(1 point)** Est-ce que la décision d'offrir un bouquet était la bonne, finalement ?

A priori et dans ce réseau, $P(A = a \mid do(T = C)) = P(A = a \mid T = C) = 0{+}0.3{+}0.2 = 0.5$ et $P(A = a \mid do(T = C)) = P(A = a \mid T = B) = 0.5{+}0{+}0.2 = 0.7$. Ce n'était pas le bon choix donc.

Exercice 2 (5 points)

Soit X_i, $i \in \{1, 2, 3, 4, 5\}$, des variables ayant toutes pour domaine $\{1, 2, 3, 4\}$. On considère l'équation :

$$\min_{x_1,x_2,x_3} \left[x_1 + 0,2x_2 - \log((x_2 + x_3) \times x_1) + \max_{x_4,x_5} \left[x_1x_4 + x_4x_3 + x_4^2x_5 \right] \right]. \qquad (1)$$

Q 2.1 **(3 points)** En observant que $\log ab = \log a + \log b$, écrivez la séquence de calculs permettant de résoudre l'équation (1) avec la séquence d'élimination x_5, x_4, x_2, x_1, x_3. Vous effectuerez ces calculs et donnerez la solution du l'équation (1). Vous pourrez vous aider du tableau ci-dessous qui fournit les logarithmes de tous les entiers de 1 à 32 (par exemple, $\log 25$ se trouve à l'intersection de la colonne 1 et de la ligne 24 et vaut donc 3,218876).

	1	2	3	4	5	6	7	8
0	0.000000	0.693147	1.098612	1.386294	1.609438	1.791759	1.945910	2.079442
8	2.197225	2.302585	2.397895	2.484907	2.564949	2.639057	2.708050	2.772589
16	2.833213	2.890372	2.944439	2.995732	3.044522	3.091042	3.135494	3.178054
24	3.218876	3.258097	3.295837	3.332205	3.367296	3.401197	3.433987	3.465736

L'équation (1) s'écrit aussi :

$$\min_{x_1,x_2,x_3} \left[x_1 + 0,2x_2 - \log(x_2 + x_3) - \log(x_1) + \max_{x_4,x_5} \left[x_1x_4 + x_4x_3 + x_4^2x_5 \right] \right].$$

Lorsque l'on supprime x_5 de l'équation, on doit calculer $\phi_5(x_4) = \max_{x_5} x_4^2x_5 = 4x_4^2$. Quelle que soit la valeur de x_4, $\phi_5(x_4)$ est obtenu pour $x_5 = 4$.
Ensuite, à la suppression de x_4 on calcule $\phi_4(x_1, x_3) = \max_{x_4} x_1x_4 + x_4x_3 + \phi_5(x_4) = 4(x_1 + x_3) + 64$ et, là encore, quelles que soient les valeurs de x_1 et x_3, l'optimum est

obtenu pour $x_4 = 4$. L'équation (1) est maintenant équivalente à :

$$\min_{x_1,x_2,x_3} \left[x_1 + 0,2x_2 - \log(x_2 + x_3) - \log(x_1) + \phi_4(x_1, x_3)\right].$$

Éliminons x_2 : il y a 2 termes dépendant de x_2 : $0,2x_2 - \log(x_2 + x_3)$. On crée donc une nouvelle fonction $\phi_2(x_3) = \min_{x_2} [0,2x_2 - \log(x_2 + x_3)]$. Pour calculer le min sans avoir à examiner tous les couples (x_2, x_3) possibles, dérivons la fonction à minimiser : $f(x_2) = 0,2x_2 - \log(x_2 + x_3)$, $f'(x_2) = 0,2 - \frac{1}{x_2+x_3}$. Par conséquent, $f'(x_2) \geq 0 \Longleftrightarrow x_2 \geq 5 - 1,5x_3$. On obtient donc, pour $x_3 = 1$, un minimum de f en $x_2 = 3,5$. Le min que l'on recherche est alors soit $x_2 = 3$, soit $x_2 = 4$. En calculant $f(3)$ et $f(4)$, on s'aperçoit que $f(4) < f(3)$. C'est donc notre minimum. Lorsque $x_3 = 2$, le min de f est obtenu en $x_2 = 2$. Lorsque $x_3 = 3$, le min de f est obtenu en $x_2 = 0,5$. En comparant $f(0)$ et $f(1)$, on s'aperçoit que le minimum que l'on recherche est obtenu pour $x_2 = 1$. Enfin, lorsque $x_3 = 4$, le min de f est obtenu en $x_2 = -1$. Donc le min que l'on recherche est obtenu pour $x_2 = 0$. En résumé, on obtient pour la fonction $\phi_2(x_3)$:

$$\phi_2(x_3) = \begin{cases} -0.809438 \text{ pour } x_3 = 1 & (\text{Argmax : } x_2 = 4) \\ -1.009438 \text{ pour } x_3 = 2 & (\text{Argmax : } x_2 = 3) \\ -1.209438 \text{ pour } x_3 = 3 & (\text{Argmax : } x_2 = 2) \\ -1.409438 \text{ pour } x_3 = 4 & (\text{Argmax : } x_2 = 1) \end{cases}$$

On doit donc maintenant résoudre :

$$\min_{x_1,x_3} \left[x_1 + \phi_2(x_3) - \log(x_1) + \phi_4(x_1,x_3)\right].$$

Donc l'élimination de x_1 consiste à calculer $\phi_1(x_3) = \min_{x_1} x_1 - \log(x_1) + \phi_4(x_1,x_3) = \min_{x_1} x_1 - \log(x_1) + 4(x_1 + x_3) + 64$. Le min est clairement obtenu pour $x_1 = 1$, quelle que soit la valeur de x_3. Par conséquent, $\phi_1(x_3) = 4x_3 + 69$. Enfin, l'élimination de x_3 consiste à calculer $\phi_3(x_3) = \min_{x_3} \phi_1(x_3) + \phi_2(x_3)$. Or,

$$\phi_1(x_3) + \phi_2(x_3) = \begin{cases} 72.190562 \text{ pour } x_3 = 1 \\ 75.990562 \text{ pour } x_3 = 2 \\ 79.790562 \text{ pour } x_3 = 3 \\ 83.590562 \text{ pour } x_3 = 4 \end{cases}$$

Donc, le minimum est obtenu pour $\widehat{x_3} = 1$. Lorsque $\widehat{x_3} = 1$, d'après ϕ_2, $\widehat{x_2} = 4$. De plus, d'après les autres ϕ_i, $\widehat{x_5} = 4$, $\widehat{x_4} = 4$, $\widehat{x_1} = 1$. Donc le quintuplet solution de l'équation (1) est $(1,4,1,4,4)$.

Q 2.2 **(1 point)** Tracez l'arbre d'*élimination* correspondant aux calculs ci-dessus. Uniquement pour la phase de collecte, vous indiquerez à côté des séparateurs les opérateurs utilisés pour passer de la clique voisine au séparateur et du séparateur à l'autre clique. Par exemple, sur le graphe de la figure 1, on passe de la clique XYZ au séparateur YZ en calculant un min sur X, et on ajoute la table du séparateur à celle de la clique YZT.

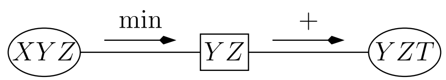
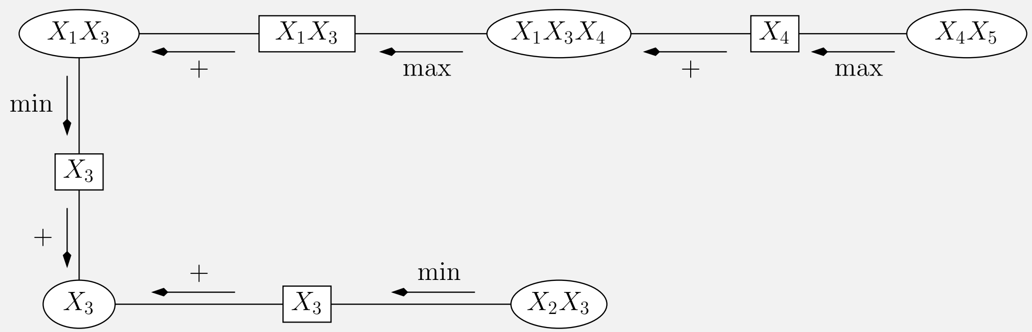
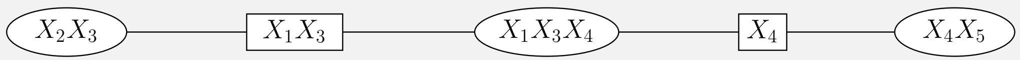


FIG. 1 – Opérateurs utilisés pendant la phase de collecte.



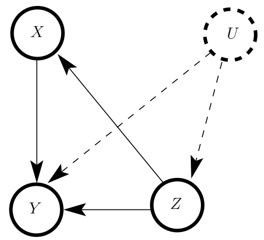
Q 2.3 **(1 point)** Tracez l'arbre de jonction correspondant.



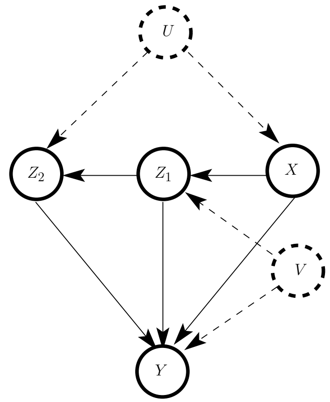
Exercice 3 (6 Points) — *Identification*

Dans les graphes suivants, peut-on calculer $P(y \mid do(x))$? Si oui, donnez son expression.

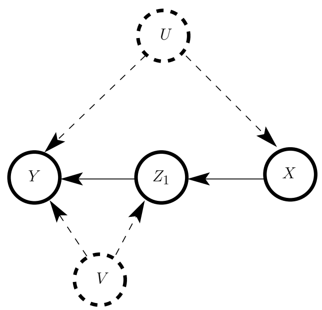
Q 3.1 (2 points)



Q 3.2 (2 points)



Q 3.3 (2 points)



Exercice 4 (10 points)

Dans cet exercice, on considère des graphes non orientés $G = (\mathcal{V}, \mathcal{E})$, où $\mathcal{V} = \{X_1, \dots, X_n\}$ et $\mathcal{E} \subseteq \{(X_i, X_j) : i, j \in \{1, \dots, n\}, i \neq j\}$. $\sigma : \{1, \dots, n\} \mapsto \{1, \dots, n\}$ représente une permutation.

Q 4.1 **(3 points)** Soit l'algorithme d'élimination :

Fonction Elim (G, σ)
- $\mathcal{E}' \longleftarrow \mathcal{E}$
- **pour** i variant de 1 à n **faire**
  - **pour** X_j, X_k adjacents à $X_{\sigma(i)}$ dans G **faire**
    - **si** $(X_j, X_k) \notin \mathcal{E}$ **alors** $\mathcal{E} \longleftarrow \mathcal{E} \cup \{(X_j, X_k)\}$; $\mathcal{E}' \longleftarrow \mathcal{E}' \cup \{(X_j, X_k)\}$ **finsi**
  - **fait**
  - supprimer toutes les arêtes adjacentes à $X_{\sigma(i)}$ dans $\mathcal{E}$
- **fait**
- renvoyer le graphe $G' = (\mathcal{V}, \mathcal{E}')$

On appelle cycle de G' un ensemble de nœuds $\{X_{i_1}, \dots, X_{i_p}\}$ tel que : i) pour tout $k \in \{1, \dots, p-1\}$, $(X_{i_k}, X_{i_{k+1}}) \in \mathcal{E}'$; ii) $(X_{i_p}, X_{i_1}) \in \mathcal{E}'$; iii) il n'existe pas $j, k \in \{1, \dots, p\}$, $j \neq k$, tels que $X_{i_j} = X_{i_k}$. La longueur d'un cycle est le nombre d'arêtes qui le constitue, autrement dit, la longueur du cycle $\{X_{i_1}, \dots, X_{i_p}\}$ est $p + 1$. Montrez que le graphe G' renvoyé par la fonction **Elim** est triangulé, i.e., dans tout cycle de longueur 4 ou plus, il existe deux nœuds X, Y non voisins dans le cycle tels que $(X, Y) \in \mathcal{E}'$.

Dans un cycle, on appelle une *corde* un couple de nœuds X, Y non voisins dans le cycle tels que $(X, Y) \in \mathcal{E}'$. Un cycle est *cordal* s'il contient au moins une corde.
Démonstration par récurrence : Dans le graphe G' renvoyé, il ne peut exister de cycle non cordal de longueur 4 ou plus passant par $X_{\sigma(1)}$. En effet, dans un tel cycle, pour tout couple X, Y de voisins de $X_{\sigma(1)}$, X ne peut être égal à Y (d'après la propriété iii) des cycles) et X ne peut être voisin de Y dans le cycle puisque ce dernier est de longueur 4 ou plus. Or, $(X, Y) \in \mathcal{E}'$ d'après la ligne 04 de l'algorithme **Elim**. Donc (X, Y) est une corde. S'il existe un cycle non cordal de longueur 4 ou plus dans G', celui-ci ne peut donc passer par $X_{\sigma(1)}$. Par conséquent, en supprimant $X_{\sigma(1)}$ ainsi que ses arêtes adjacentes de G', on ne peut supprimer ce cycle. Dans le graphe résultant, on peut appliquer la même démonstration pour $X_{\sigma(2)}$. Par récurrence, on peut supprimer tous les nœuds de G' et montrer ainsi qu'aucun cycle de longueur 4 ou plus non cordal ne peut passer par eux.
Conséquence : le graphe G' est triangulé.

Q 4.2 **(2 points)** On appelle graphe markovien associé à un arbre de jonction $\mathcal{J}$ le graphe $G = (\mathcal{V}, \mathcal{E})$ tel que $\mathcal{V}$ est l'ensemble des variables contenues dans les cliques de $\mathcal{J}$ et $\mathcal{E}$ est l'ensemble des arêtes (X_i, X_j) telles qu'il existe une clique dans $\mathcal{J}$ contenant X_i et X_j. Montrez que tout graphe markovien associé à un arbre de jonction est triangulé.

Soit $\mathcal{J}$ un arbre de jonction et G le graphe markovien associé. On montre la propriété par récurrence sur le nombre de cliques de $\mathcal{J}$.
Si $\mathcal{J}$ ne contient qu'une seule clique, G est un graphe complet et, par conséquent, il est triangulé. Sinon, puisque $\mathcal{J}$ est un arbre, il existe une clique, appelons-la C_1, ayant exactement une clique voisine (appelons-la C'). Soit $S = C_1 \cap C'$ le séparateur entre C_1 et C'. Les nœuds de $D_1 = C_1 \backslash C'$ ne peuvent appartenir qu'à la clique C_1 (propriété d'intersection courante). Or, C_1 est une clique, donc tous les voisins des nœuds de D_1 sont reliés entre eux. Par conséquent, il ne peut exister de cycle de longueur 4 ou plus non cordal passant par les nœuds de D_1 (puisque lesdits voisins forment des cordes). Les autres nœuds de la clique C_1 appartiennent à S, et donc a fortiori à C'. Donc, si l'on supprime la clique C_1 de $\mathcal{J}$, le graphe markovien associé à ce nouvel arbre de jonction (appelons-le $\mathcal{J}^1$) est le graphe G^1 obtenu en supprimant de G les nœuds de D_1 ainsi que leurs arêtes adjacentes. S'il existe un cycle non cordal de longueur 4 ou plus dans G, alors il existe aussi dans G^1. On peut donc appliquer à nouveau la démonstration ci-dessus avec $\mathcal{J}^1$. Le nombre de cliques de $\mathcal{J}^1$ etant égal à celui de $\mathcal{J}$ moins 1, la récurrence s'arrêtera lorsqu'il n'y aura plus qu'une seule clique, cas où l'on a déjà vu que le graphe markovien était triangulé.

Q 4.3 Considérons l'algorithme d'élimination orienté suivant :

Fonction DirectElim $(G = (\mathcal{V}, \mathcal{E}), \sigma)$

$\mathcal{A} \longleftarrow \emptyset$
pour i variant de 1 à n **faire**
  pour tous X_k adjacents à $X_{\sigma(i)}$ dans G **faire** $\mathcal{A} \longleftarrow \mathcal{A} \cup \left\{\overrightarrow{X_{\sigma(i)} X_k}\right\}$ **fait**
  pour tous X_j, X_k adjacents à $X_{\sigma(i)}$ dans G **faire**
    si $(X_j, X_k) \notin \mathcal{E}$ **alors** $\mathcal{E} \longleftarrow \mathcal{E} \cup \{(X_j, X_k)\}$; **finsi**
  fait
  supprimer toutes les arêtes adajacentes à $X_{\sigma(i)}$ dans $\mathcal{E}$
fait
renvoyer le graphe orienté $G' = (\mathcal{V}, \mathcal{A})$

Le graphe $G' = (\mathcal{V}, \mathcal{A})$ renvoyé par **DirectElim** est un graphe orienté tel que, si l'on enlève les orientations des arcs de $\mathcal{A}$, le graphe résultant est triangulé.

$\boxed{\textbf{Q 4.3.1}}$ **(1 point)** Dessinez le graphe G' obtenu à l'issue de l'application de **DirectElim** sur le graphe de la figure 2 et la séquence d'élimination $\sigma = \{A, C, B, E, D, F, G\}$:

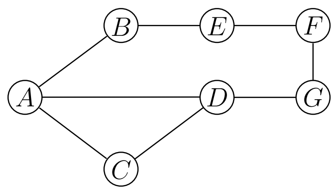
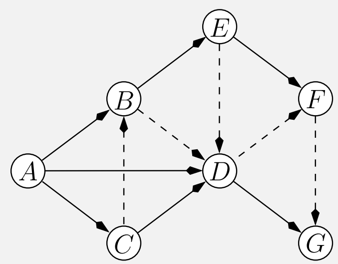


FIG. 2 – Un graphe non orienté à trianguler.



Les arcs en pointillés sont les fill-ins rajoutés par l'algorithme d'élimination.

Q 4.3.2 **(2 points)** Montrer que, pour tout arc $\overrightarrow{XY} \in \mathcal{A}$, X est éliminé avant Y dans l'algorithme **DirectElim**.

D'après l'algorithme, lorsqu'un arc $\overrightarrow{XY}$ est rajouté à $\mathcal{A}$, cela signifie qu'on est en train d'éliminer le nœud X (cf. ligne 03). Si le nœud Y avait été éliminé avant X alors ce nœud n'aurait pu créer l'arc $\overrightarrow{XY}$ (il aurait éventuellement créé un arc $\overrightarrow{YX}$) sur la ligne 03. Or, à la fin de l'élimination de Y, sur la ligne 07, Y et ses arêtes adjacentes dans G sont éliminées. Par conséquent, il ne serait pas possible de créer ultérieurement l'arc $\overrightarrow{XY}$. En conséquence, X doit être éliminé avant Y.

Q 4.3.3 **(2 points)** Dans le graphe G' de la question 4.3.1, on veut supprimer l'arc $\overrightarrow{FG}$ de $\mathcal{A}$ de telle sorte que le graphe ainsi obtenu reste triangulé. Pour cela, à tout arc $\overrightarrow{YZ}$ de $\mathcal{A}$, on associe un nombre entier $x_{YZ} \in \{0, 1\}$ ayant pour signification $x_{YZ} = 1 \Longleftrightarrow$ l'arc

$\overrightarrow{YZ}$ est supprimé. On impose donc que $x_{FG} = 1$. Soit le programme linéaire (Σ) défini par :

$$\begin{cases} x_{YZ} \leq x_{XY} + x_{XZ} \ \forall X, Y, Z \text{ tels que } \overrightarrow{XY}, \overrightarrow{XZ}, \overrightarrow{YZ} \in \mathcal{A} \\ x_{FG} = 1 \\ x_{YZ} \in \{0,1\} \ \forall \overrightarrow{YZ} \in \mathcal{A} \end{cases} \tag{Σ}$$

Montrer que, pour toute solution du programme (Σ), le graphe $G'' = (\mathcal{V}, \mathcal{A}'')$ constitué uniquement des arcs pour lesquels les x_{YZ} valent 0 est un graphe triangulé. Suggestion : démonstration par récurrence sur les nœuds éliminés, selon leur ordre d'élimination.

Supposons que les $x_{X,Y}$ sont solutions de (Σ). Si l'on élimine le nœud A en premier, alors les équations du type $x_{YZ} \leq x_{AY} + x_{AZ}$ signifient simplement que l'arc $\overrightarrow{YZ}$ ne peut être éliminé que si au moins un des deux arcs $\overrightarrow{AY}$ ou $\overrightarrow{AZ}$ l'est aussi. Considérons maintenant un cycle de longueur 4 ou plus passant par A dans G''. Le cycle passe donc aussi par 2 voisins de A dans G'', appelons-les Y et Z. Si $x_{YZ} = 1$, alors soit $x_{AY} = 1$, soit $x_{AZ} = 1$, soit les 2. Dans tous les cas, cela signifie qu'au moins un des 2 nœuds Y ou Z n'est pas voisin de A, d'où contradiction puisque Y et Z sont voisins de A dans G''. Par conséquent, on a nécessairement $x_{YZ} = 0$ et donc l'arc $\overrightarrow{YZ}$ appartient bien à $\mathcal{A}''$ et le cycle passant par A possède une corde : l'arc $\overrightarrow{YZ}$. S'il existe un cycle de longueur 4 ou plus non cordal dans G'', alors celui-ci ne passe par par A. On peut donc supprimer A

et ses arcs adjacents et itérer la démonstration ci-dessus sur les autres nœuds éliminés, selon leur ordre d'élimination.