Examen sur les Réseaux Bayésiens

mercredi 19 avril 2017

Durée de l'épreuve: 3h. Notes de cours interdites, calculatrice autorisée.
Barême approximatif donné à titre purement indicatif et susceptible de changements.
Les deux exercices peuvent être traités indépendamment l'un de l'autre.

Exercice 1 (8 points)

On s'intéresse à la génétique d'une famille de $n = 8$ individus. Pour tout $i \in \{1, \dots, n\}$ on désigne par $X_i \in \{\mathtt{00}, \mathtt{01}, \mathtt{10}, \mathtt{11}\}$ le génotype de l'individu i. On admet que la loi de $X = (X_i)_i$ est définie par le réseau bayésien de la figure 1. On cherche à déterminer la distribution de X sachant l'évidence $\mathrm{ev} = \{X_1 = X_8 = \mathtt{11}\}$.

1) [**1 point**] Quelle est la relation familiale entre les individus 1 et 8 ? On admet que 2 et 5 sont des femmes, qu'en déduisez-vous sur le sexe des autres individus ?

2) [**1 points**] Ecrire $\mathbb{P}(X)$ sous la forme d'un produit de probabilité conditionnelles (sans détailler ces quantités) et vérifier que $\sum_X \mathbb{P}(X) = 1$. Ecrire $\mathbb{P}(X, \mathrm{ev})$ sous la forme d'un produit de potentiels.

3) [**3 points**] Proposer un *junction tree* ayant une *tree-width* inférieure à 4 pour le réseau bayésien. On vérifiera soigneusement que le *junction tree* proposé vérifie bien toutes les propriétés nécessaires.

4) [**1 point**] Quelle est le nombre de configuration à considérer pour calculer $\mathbb{P}(\mathrm{ev})$: a) par la force brute; b) à l'aide de votre *junction tree*.

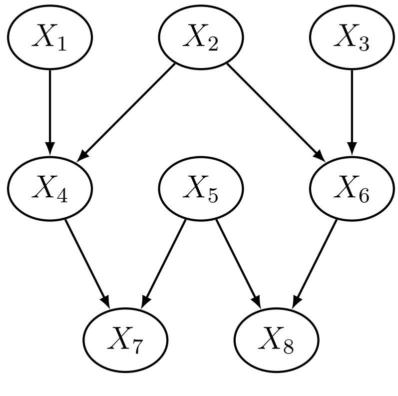

Figure 1: Réseau bayésien de l'exercice 1.

5) [**2 point**] Choisir une arête $C_i - C_j$ de votre *junction tree* et donner, sous la forme d'une somme-produit, l'expression des messages $M_{i \to j}$ et $M_{i \to j}$. Donner egalement une expression de ces deux messages à l'aide des autres messages et des potentiels du réseau bayésien.

Exercice 2 (12 points)

On s'intéresse dans cet exercice au processus autoregressif d'ordre 1 défini pour $i = 1, \ldots, n$ par

$$X_i = \eta X_{i-1} + e_i$$

avec $X_0 = a \in \mathbb{R}$, $\eta \in]0, 1[$ et où les e_i sont iid de loi gaussienne $\mathcal{N}(\delta, \varepsilon^2)$. Pour un certain $b \in \mathbb{R}$, on cherche à étudier la distribution $\mathbb{P}(X_1, \ldots, X_n | X_n = b)$.

1) [**2 points**] Décrire le problème sous la forme d'un réseau bayésien dont on donnera tous les potentiels (on pourra par exemple noter $\texttt{dnorm}(\cdot, \mu, \sigma)$ la densité d'une gaussienne $\mathcal{N}(\mu, \sigma^2)$).

2) [**1 points**] Expliquer pourquoi on introduit alors les quantités *forward* et *backward* définies pour tout $i = 1, \ldots, n$ et tout $x \in \mathbb{R}$ par

$$F_i(x) = \mathbb{P}(X_i = x) \quad \text{et} \quad B_i(x) = \mathbb{P}(X_n = b | \ X_i = x).$$

où $\mathbb{P}(X_i = \cdot)$ désigne la densité marginale de X_i et où $\mathbb{P}(X_n = \cdot | \ X_i = x)$ designe la densité de X_n conditionnellement à $\{X_i = x\}$.

3) [**3 points**] prouver que

$$F_i(x) B_i(x) = \mathbb{P}(X_i = x, X_n = b) \text{ pour } i = 1, \ldots, n$$

et

$$F_{i-1}(x) \texttt{dnorm}(y, \eta x + \delta, \varepsilon) B_i(y) = \mathbb{P}(X_{i-1} = x, X_i = y, X_n = b) \text{ pour } i = 2, \ldots, n-1$$

où les probabilités désignent encore (abusivement) les densités jointes correspondantes.

4) [**2 points**] Déduire de la question 3 des formules récursives permettant de calculer les densités F_i et B_i pour tout i.

5) [**2 points**] Déduire de la question 3 une expression de la densité conditionnelle $\mathbb{P}(X_i = y | X_{i-1} = x, X_n = b)$. Décrire un algorithme permettant de générer une réalisation de $X_1, \ldots, X_n$ conditionnellement à $\{X_n = b\}$

6) [**2 points**] On s'intéresse maintenant à la variable aléatoire $N(c) = \sum_{i=1}^{n} \mathbf{1}_{X_i > c}$ où $c \in \mathbb{R}$, décrire (sans détailler les calculs) une approche permettant d'obtenir la distribution $\mathbb{P}(N(c) | \ X_n = b)$.