Examen du cours sur les Modèles Graphiques Probabilistes

Christophe Ambroise

5/28/2020

L'examen à distance se déroule de 9h à 12h.

- 9 heures: téléchargement de l'examen. Vous devez me confirmer que le fichier a bien été téléchargé en m'envoyant un mail à christophe.ambroise@univ-evry.fr.
- 12 heures (au plus tard): envoi de la solution sous forme du fichier R markdown complété et ou de photos de vos feuilles pour les calculs à mon adresse mail christophe.ambroise@univ-evry.fr avec pour objet examen-MG.

Chacun d'entre vous a un profil (A, B ou C). Dans le cas où les questions d'un exercice sont notées (a, b, c) vous répondrez à celle correspondant à votre profil.

En cas de problème pendant l'examen vous pouvez me retrouver

- sur ecampus: https://eu.bbcollab.com/guest/16ed88329c8648c7b27c9664a5c8a809
- ou par téléphone 06 70 76 82 60

---

numéro	nom	profil
20177039	ABRAICH	C
20171370	AMOYALA	A
20182593	BA	B
20144090	BECHET	C
20195812	BERRAHO	A
20194379	BOUBEKEUR	B
20156921	BOUCHOUAT	C
20195857	BOUIZAGUEN	A
20193934	CHEN	B
20196128	CHOKKI	C
20194496	DAVIDAS-ROCH	A
20195268	DIALLO	B
20193490	EL-ISSATI	C
20182405	GHOUL	A
20142071	GUEDJ	B
20191684	HOANG	C
20195584	KAINA	A
20194863	LBAHY	B
20194462	MAHRAOUI	C
20164685	MARCOUX-PEPIN	A
20177989	MARDOC	B
20195999	NGREMMADJI	C
20193377	NGUYEN	A
20195420	NIANG	B
20193583	TCHOUACHEU	C
20182332	TOUNSI	A
20194501	TOURE	B

Réseau de gènes (7 points)

Un réseau de régulation génique (ou génétique) est un ensemble de régulateurs moléculaires qui interagissent entre eux et avec d'autres substances dans la cellule pour régir les niveaux d'expression génique. La Figure 1 représente un réseau de 4 gènes A, B, C, D. Chaque gène est exprimé (1) ou pas (0).

1. Donner la factorisation de la densité jointe des 4 gènes.
2. Lister les indépendances conditionnelles.
3. Donner les paramètres du modèle.
4. Ecrire sur le papier un code R qui produit 100 réalisations des 4 variables A, B, C, D (sans utiliser de package).
5. a. Donner la couverture de Markov de A.
   b. Donner la couverture de Markov de B.
   c. Donner la couverture de Markov de C.
6. Donner l'expression des estimateurs du maximum de vraisemblance des paramètres du modèle.

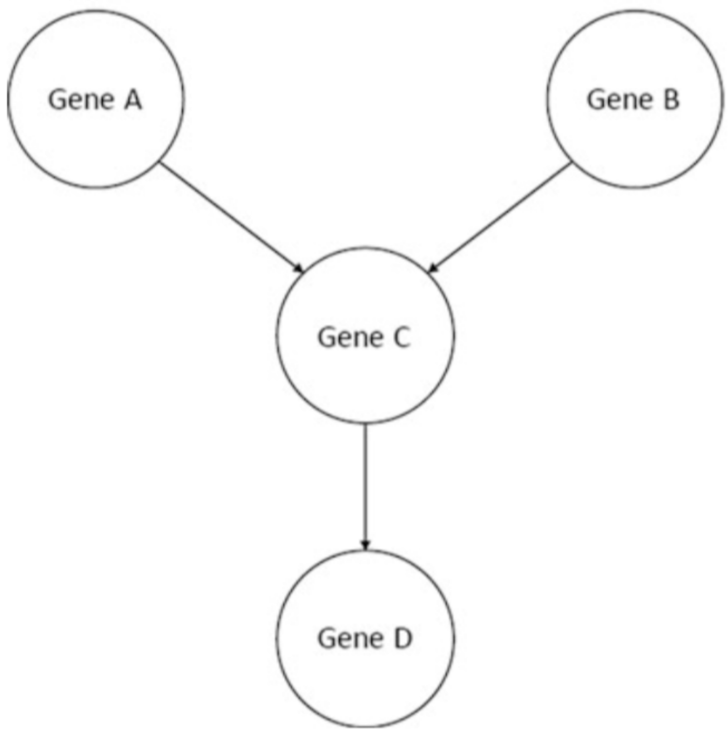

Figure 1: Réseau de 4 gènes et tables de probabilités associées

Indépendance conditionnelle (3 points)

On considère des jeux de variables discrètes disjointes: $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$. Mopntrer que s'il existe deux fonctions F et G telles que $P(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = F(\boldsymbol{X}, \boldsymbol{Z})G(\boldsymbol{Y}, \boldsymbol{Z})$ alors $\boldsymbol{X}$ est indépendant de $\boldsymbol{Y}$ sachant $\boldsymbol{Z}$.

Indépendance conditionnelle (2 points)

Pour le graphe de Markov ci-dessous (Figure 2), listez toutes les relations d'indépendance conditionnelle implicites et trouvez les cliques maximales.

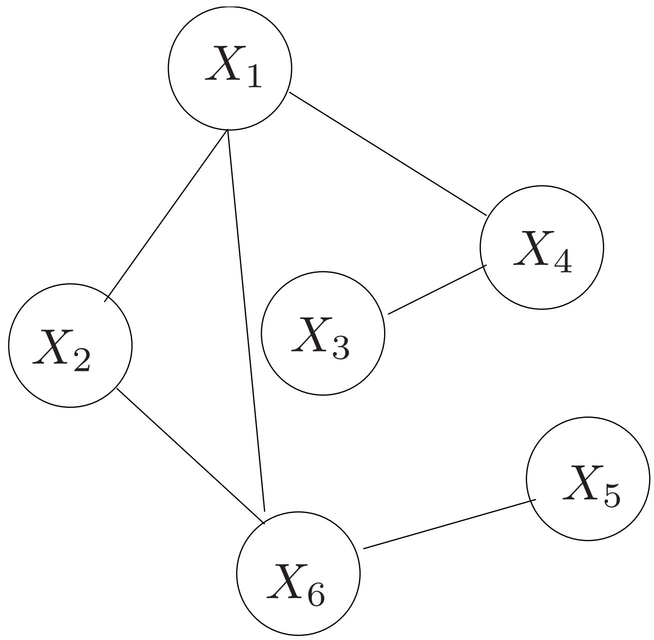

Figure 2: Graphe de Markov

Des indépendances vers le graphe (1 points)

Soit les variables aléatoires X_1, X_2, X_3, X_4. Dans chacun des cas suivant dessiner un graphe qui respecte les indépendances données.

a. $X_1 \perp\!\!\!\perp X_3 \mid X_2$, and $X_2 \perp\!\!\!\perp X_4 \mid X_3$.
b. $X_1 \perp\!\!\!\perp X_4 | X_2, X_3$ and $X_2 \perp\!\!\!\perp X_4 \mid X_1, X_3$.
c. $X_1 \perp\!\!\!\perp X_4 \mid X_2, X_3, X_1 \perp\!\!\!\perp X_3 | X_2, X_4$ et $X_3 \perp\!\!\!\perp X_4 \mid X_1, X_2$.

Modèle d'Ising avec variable binaire (7 points)

En notant la variable **binaire** au noeud j par X_j, un modèle d'Ising sur un graphe G avec p noeuds a pour loi jointe

$$p(X, \theta) = \exp\left(\sum_{j \sim k} \theta X_j X_k - \Phi(\theta)\right)$$

avec$X \in \mathcal{X} = \{0,1\}^p$, $j \sim k$ signifiant que les noeuds j et k définissent une arête de G et

$$\Phi(\theta) = \log\Big(\sum_{x \in \mathcal{X}} \exp\big(\sum_{j \sim k} \theta x_j x_k\big)\Big)$$

1. Caculer $Pr(X_j = 1 | X_{\setminus j} = x_{\setminus j})$ où $X_{\setminus j}$ et $Pr(X_j = 0 | X_{\setminus j} = x_{\setminus j})$ où $X_{\setminus j}$ où $x_{\setminus j}$représente tous les noeuds du graphe sauf le noeud j.
2. Supposons que l'on observe une réalisation de la loi jointe du modèle d'Ising. En utilisant la question 1, écrire la somme des p log probabilités conditionnelles des p noeuds.

3. Quels liens et différences entre un modèle de régression logistique et le modèle d'Ising considéré.
4. Ecrire un pseudo-code "R" ou python, qui estime θ en utilisant les questions précédentes.
5. Calculer le gradient par rapport à θ de la log-vraisemblance.
6. Quelle difficulté pratique voyez vous dans l'utilisation d'une descente de gradient pour l'estimation de θ?