

IFT3335 - TP2

Réseau bayésien et inférence

Ce TP est à réaliser en groupe de 2 personnes (ou seul).

Date de remise : 21 novembre, avant 17:30.

Vous devez seulement remettre votre rapport de 5 pages au plus.

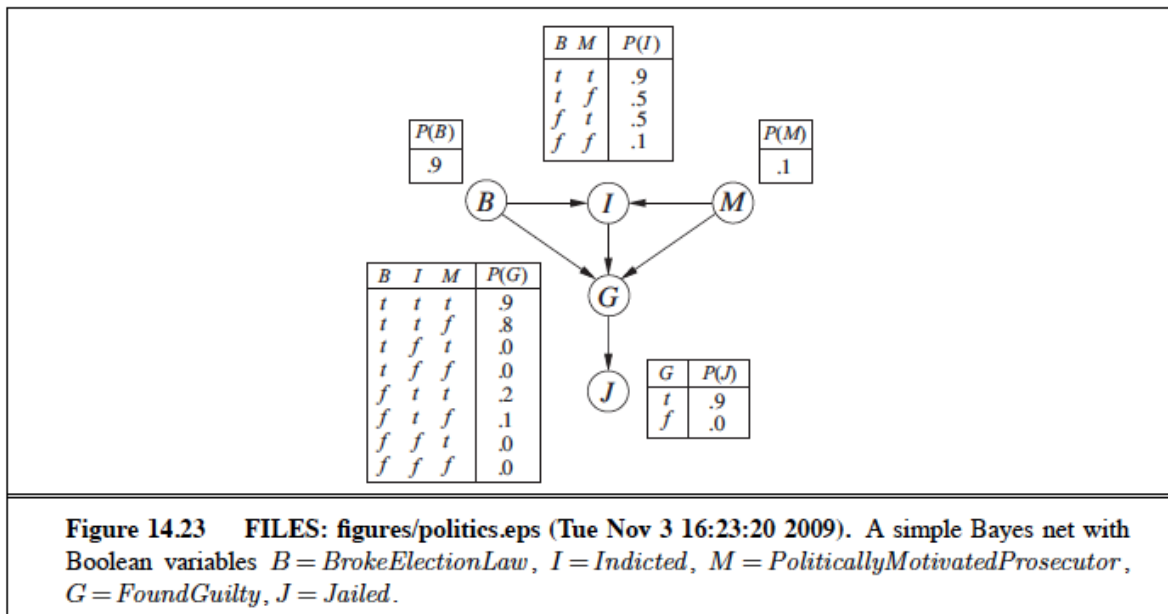
Ce TP compte pour 10% de la note globale.

Ce TP a pour but de pratiquer l'inférence dans un réseau bayésien – un type de modèle graphique largement utilisé pour le traitement de l'incertitude. Outre la méthode traditionnelle d'inférence qui utilise la théorie de probabilité (les formules) pour faire le calcul exact, il y a une autre famille de méthodes basées sur l'échantillonnage. Ce TP vise à comparer ces deux méthodes, et surtout de pratiquer la deuxième famille de méthodes.

Le package de programmes qui accompagne le livre de référence contient des implantations des algorithmes d'inférence. Dans ce TP, on vous demande d'utiliser ces programmes pour faire l'inférence dans 2 problèmes.

Problème 1.

Considérons le même problème que vous avez traité dans Exercice 3, qui est modélisé par le réseau bayésien suivant :



On vous demande de déterminer les probabilités suivantes en utilisant différentes méthodes :

1. la probabilité que quelqu'un viole la loi électorale (*BrokeElectionLaw*) est jugé coupable (*FoundGuilty*), i.e. $P(g|b)$;
2. la probabilité conjointe que quelqu'un viole la loi électorale (*BrokeElectionLaw*), fait face à un procureur avec des motivations politique (*PoliticallyMotivatedProsecutor*), et n'est pas mis en prison (*Jailed*), i.e. $P(g, b, m, \neg j)$.

Pour ce problème, on vous demande d'utiliser les méthodes d'inférence suivantes :

- Calculer sur papier la probabilité exacte en utilisant la formule que vous voulez.
- Utiliser le programme d'inférence par énumération (*Enumeration-Ask*) ;
- Utiliser le programme d'inférence par échantillonnage simple (*Prior-Sample*) ;
- Utiliser le programme d'inférence par échantillonnage Gibbs (*Gibbs-Ask*).

Observez si les valeurs obtenues avec les programmes sont identiques ou proches de celles que vous obtenez dans votre calcul sur papier.

Dans les cas d'échantillonnage, le nombre d'échantillons joue un rôle important : plus on utilise d'échantillons, plus la valeur qu'on obtient sera proche de la vraie valeur exacte. Pour observer ceci, faites varier le nombre d'échantillons, par exemple, à 10, 100, 1000, et observez l'impact de ce nombre sur les valeurs.

Vous devez discuter de vos observations dans votre rapport.

Problème 2.

Nous avons 3 équipes de football A, B et C. Chaque équipe doit s'affronter avec chaque autre équipe une fois. Une équipe contre une autre équipe peut avoir le résultat gagné (G), nul (N) ou perdu (P). On suppose que chaque équipe a une qualité fixe, qu'on ne connaît pas. On représente cette qualité par un nombre entier, de 0 (le moins bon) à 3 (le meilleur). Le résultat de chaque match est déterminé de façon probabiliste selon la qualité des deux équipes. Pour cet exercice, on suppose la table de probabilité suivante :

X	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
Y	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
G	0.3	0.2	0.1	0.05	0.5	0.3	0.2	0.1	0.7	0.5	0.3	0.2	0.9	0.7	0.5	0.3
N	0.4	0.3	0.2	0.05	0.3	0.4	0.3	0.2	0.2	0.3	0.4	0.3	0.05	0.2	0.3	0.4
P	0.3	0.5	0.7	0.9	0.2	0.3	0.5	0.7	0.1	0.2	0.3	0.5	0.05	0.1	0.2	0.3

où X et Y représentent deux équipes, avec leurs qualités respectives, et les nombre dans chaque colonne représentent la probabilité pour X de gagner, faire un match nul ou perdre.

- Dessinez un réseau bayésien pour ce problème.
- Supposons qu'on connaît les résultat de deux premiers matchs : A a gagné contre B, et B et C ont eu un match nul. Utilisez le calcul exact (selon les formules) pour déterminer la distribution de probabilité pour chaque résultat possible pour le dernier match, A contre C.
- Utilisez le programme avec l'échantillonnage Gibbs pour obtenir ces probabilités. Observez à quelle vitesse ces probabilités convergent, selon le nombre d'échantillons utilisés. Pour observer cela, calculez les probabilités en utilisant des nombres d'échantillons croissants, jusqu'à une convergence (des probabilités assez proches des valeurs exactes). Discutez sur la convergence dans votre rapport.
- Supposons maintenant que nous avons 5 équipes, et que nous connaissons les résultats de tous les matchs, sauf le dernier. Supposez les résultats de ces matchs connus comme vous voulez. Estimez la probabilité de chaque résultat possible du dernier match avec échantillonnage Gibbs. À quelle vitesse ce processus se converge ?

Rapport :

Dans votre rapport, vous devez décrire les expériences que vous faites, les résultats obtenus, et surtout une analyse sur les résultats. Décrivez ce que vous observez d'intéressant dans ces expériences.

Votre rapport peut être remis des façons suivantes :

- sur papier aux démonstrateurs,
- par le programme remise : remise ift3335 TP2 <votre-rapport>
- ou par courriel à dift3335@iro.umontreal.ca

Barème :

- Problème 1 : 40%
 - o 10% pour chaque sous-problème a, b, c et d.
- Problème 2 : 40%
 - o 10% pour chaque sous-problème a, b, c et d.
- Rapport : 20%

Pour le rapport, on vous suggère d'inclure au moins les éléments suivants :

1. Description brève de ce que vous avez fait pour chaque problème ;
2. Vos expériences, observations et analyses.