

# Modélisation de l'incertitude — Master 2 MIAGE IA<sup>2</sup>

## Travaux dirigés N° 3 : Réseaux bayésiens\*

Andrea G. B. Tettamanzi  
Université côte d'Azur  
andrea.tettamanzi@univ-cotedazur.fr

Année universitaire 2020/2021

### Résumé

On va construire et utiliser des réseaux bayésiens en se basant sur une bibliothèque Python, Pomegranate en l'occurrence.

## 1 Introduction

Le but de ce TP va être de vous familiariser avec les réseaux bayésiens. Pour ce faire, nous nous baserons sur l'utilisation de Pomegranate (<https://pomegranate.readthedocs.io/en/docs/index.html>), un module Python qui réalise une grande variété de modèles probabilistes, parmi lesquelles les réseaux bayésiens. Pour son installation, il suffit de taper la commande

```
pip3 install pomegranate
```

Attention : il faut que le module Cython soit déjà installé, sinon l'installation pourrait échouer. Sous Windows, il pourrait être nécessaire d'installer aussi un compilateur C++ au préalable.

## 2 Prise en main

1. Construction d'un réseau : Pour prendre en main Pomegranate, nous allons étudier la situation suivante :

Un arbre peut perdre ses feuilles pour deux raisons : soit il est malade, soit le sol est trop sec. Il y a a priori 10% de chances qu'il soit malade ; de même, la probabilité que le sol soit sec est de 10%. Les probabilités conditionnelles de perte de feuilles en fonction de la sécheresse et de la maladie sont les suivantes :

	sec		¬sec	
	malade	¬malade	malade	¬malade
perd_feuilles	0.95	0.85	0.90	0.02
¬perd_feuilles	0.05	0.15	0.10	0.98

---

\*Ce sujet reprend un sujet de travaux pratiques proposé par Matthieu Amiguet

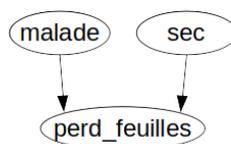


FIGURE 1 – Un réseau simple.

Nous allons donc construire dans Pomegranate le réseau de la Figure 1. Pour ce faire, nous allons commencer par importer du module Pomegranate les éléments dont nous aurons besoin :

```
from pomegranate import *
```

Ensuite, nous créons le réseau lui-même :

```
G = BayesianNetwork('BotanicalProblem')
```

L'argument est simplement le nom du réseau et n'a au fait pas grande importance. Créons maintenant les variables aléatoires de notre réseau :

```
malade = DiscreteDistribution({'Oui': 0.1, 'Non': 0.9})
sec = DiscreteDistribution({'Oui': 0.1, 'Non': 0.9})
perd_feuilles = ConditionalProbabilityTable(
[[ 'Non', 'Non', 'Non', .98],
 [ 'Non', 'Non', 'Oui', .02],
 [ 'Non', 'Oui', 'Non', .15],
 [ 'Non', 'Oui', 'Oui', .85],
 [ 'Oui', 'Non', 'Non', .1],
 [ 'Oui', 'Non', 'Oui', .9],
 [ 'Oui', 'Oui', 'Non', .05],
 [ 'Oui', 'Oui', 'Oui', .95]], [malade, sec])
```

Ensuite, il faut créer la structure du réseau, comme suit :

```
s = State(malade, name = 'malade')
d = State(sec, name = 'sec')
l = State(perd_feuilles, name = 'perd_feuilles')
G.add_states(s, d, l)
G.add_edge(s, l)
G.add_edge(d, l)
G.bake()
```

Notre réseau est prêt. Ajoutez encore la ligne suivante :

```
print(G)
```

Sauvegardez puis exécutez votre code. Vous devriez voir une description du réseau s'afficher sur la sortie standard. Notre réseau est maintenant complet ; nous allons tenter de l'utiliser pour du raisonnement.

2. Inférence bayésienne. On peut maintenant directement calculer les probabilités des valeurs de toutes les variables, avec la méthode `predict_proba`, qui prend un dictionnaire d'observations (les évidences). On commencera par un réseau sans évidences :

```
beliefs = G.predict_proba({})
beliefs = map(str, beliefs)
for state, belief in zip(G.states, beliefs):
    print(state.name, belief)
```

Exécutez votre programme ; vous devriez voir qu'en l'absence d'évidence, nos arbres ont presque 82% de chances de garder leur feuilles. Maintenant, en regardant dans le jardin, vous constatez que les arbres perdent quand même leurs feuilles :

```
beliefs = G.predict_proba({'perd_feuilles': 'Oui'})
beliefs = map(str, beliefs)
for state, belief in zip(G.states, beliefs):
    print(state.name, belief)
```

Il y a donc 47% de chances que votre sol soit sec et 49% pour que vos arbres soient malades. Vous arrosez régulièrement vos plantation et pourtant elles perdent toujours leurs feuilles ?

```
beliefs = G.predict_proba({'perd_feuilles': 'Oui', 'sec': 'Non'})
beliefs = map(str, beliefs)
for state, belief in zip(G.states, beliefs):
    print(state.name, belief)
```

Vous avez donc 83% de chances que vos arbres soient malades...

### 3 Licencié ou augmenté ?

Dans l'entreprise Bayes&Co, il y a un patron despotique qui décide périodiquement de licencier des employés et d'augmenter le salaire d'autres employés. Sa politique n'est pas très populaire car il y a autant de licenciements que d'augmentations.

Chaque fois que le patron prend une de ces décisions, elle doit être traitée par le département des ressources humaines et par la comptabilité, qui envoient ensemble une notification à l'employé (mal)heureux.

Le département des ressources humaines compte trois employés : Janet, Jim et Julia. Lors d'un licenciement, Janet s'occupe du dossier dans 30% des cas, Jim dans 35% et Julia 35%. Si par contre il s'agit d'une augmentation de salaire, les proportions passent à 20%, 20% et 60% respectivement.

À la comptabilité travaillent James et Jules. Lors d'un licenciement, James s'occupe de 37% des cas (et Jules de 63%), alors que lors d'une augmentation la proportion passe à 42% (et donc 58% pour Jules). Les employés concernés peuvent être avertis au choix par mail ou par courrier. Il suffit que les deux responsables du dossier se mettent d'accord. On constate donc que Janet et James écrivent une lettre dans 40% des cas (et donc un mail dans 60% des cas) ; mais lorsque Janet collabore avec Jules, ces proportions sont de 70% et 30% ; pour Jim et James 30% et 70% ; pour Jim et Jules 65% et 35% ; pour Julia et James 25% et 75% ; enfin pour Julia et Jules, 55% et 45%.

Sur la base de cette situation :

1. Dessinez (sur papier!) un réseau bayésien représentant cette situation.
2. En utilisant Pomegranate, écrivez un programme implémentant ce réseau.
3. Utilisez ce programme pour répondre aux questions suivantes :
  - (a) Jeremy a reçu ce matin une lettre de Janet et Jules ; quelle est la probabilité qu'il s'agisse d'une lettre de licenciement ?
  - (b) Et s'il a reçu une lettre de Jim et James ?
  - (c) Si tout ce que Jeremy sait, c'est que Julia est en charge d'un dossier le concernant, quels sont ses risques d'être licencié ?
  - (d) Jeremy a reçu un mail ce matin, mais il ne sait pas qui a traité son dossier. Quelle probabilité a-t-il d'y lire une annonce d'augmentation de salaire ?

### 4 Enquête

Un cambriolage a été constaté au siège de l'entreprise BN Tech. Une enquête préliminaire a permis de montrer que le coupable a opéré seul et qu'il s'agit forcément de l'un des 100 employés que compte l'entreprise.

Quelques jours plus tard, la police a reçu une lettre anonyme accusant Pierre Kiroul, un nouvel employé de BN Tech. La police a immédiatement relevé ses empreintes digitales et les a

comparé à celles retrouvées sur le lieu du forfait. Aucune trace correspondant à M. Kiroul n'a pu être mise en évidence. Mais selon l'inspecteur de la police scientifique chargé de l'affaire, il ne faudrait pas en tirer des conclusions hâtives. En effet, le vol a pu être effectué avec des gants. Environ 45% des cambrioleurs recourent à cette technique (alors que la probabilité de mettre des gants sans faire de cambriolage peut être considérée comme nulle, dans ce contexte). Si le voleur n'a pas mis des gants, il y a 99% de chances que nous détectons ses empreintes ; mais avec des gants, cette chance tombe à 2%. La probabilité d'observer les empreintes de quelqu'un de non-coupable sur les lieux du crime est de 5%, estime l'expert de la police.

À ce point-là de l'histoire (1), Pierre semblait presque hors de cause. Mais la police avait également trouvé sur le lieux du forfait quelques cheveux du coupable, ce qui lui a permis de faire une analyse ADN. Cette analyse a provoqué un retournement de situation : l'ADN du coupable correspondait à celui de Pierre ! Ce test est d'une grande fiabilité ; les taux de faux positifs et de faux négatifs sont tous deux à 0.1%. Autant dire que nous avons maintenant une quasi-certitude que M. Kiroul est le coupable, triomphe l'inspecteur. (2)

Pierre Kiroul a donc été accusé ; à son procès, son avocat a brillamment montré que Pierre était dans l'impossibilité de se fournir le type de gants nécessaire à la réalisation propre du crime dont on l'accusait. L'hypothèse que Pierre portait des gants doit donc être écartée. (3)

M. Kiroul va-t-il être condamné ou acquitté ?

Sur la base du palpitant thriller ci-dessus :

1. Représenter la situation sous forme de réseau bayésien (sur papier, avec les tables de probabilité).
2. En utilisant Pomegranate, écrivez un programme implémentant ce réseau.
3. Utiliser cette implémentation pour calculer la probabilité de culpabilité de Pierre Kiroul aux points (1), (2) et (3) de l'histoire ci-dessus.

Rendez votre code, vos réponses aux questions et vos observations dans un archive zippé par courriel.