Inférence bayésienne

Dans tout le chapitre, E désigne l'ensemble des issues d'une expérience aléatoire et P désigne une loi de probabilité sur E.

I - Probabilité conditionnelle :

1. Soit *A* et *B* deux événements avec $P(A)\neq 0$. On appelle probabilité conditionnelle de *B* sachant *A*, la probabilité que l'événement *B* se réalise sachant que l'événement *A* est réalisé.
   Elle est notée $P_A(B)$ ou P(B/A) et est définie par : $P_A(B)=\dfrac{P(A\cap B)}{P(A)}$

Exemples :

a) On tire une carte au hasard dans un jeu de 32 cartes.
Soit *A* l'événement "Le résultat est un pique".
Soit *B* l'événement "Le résultat est un roi".
Donc $A\cap B$ est l'événement "Le résultat est le roi de pique".

Alors : $P(A)=\dfrac{8}{32}=\dfrac{1}{4}$ et $P(A\cap B)=\dfrac{1}{32}$.
Donc la probabilité que le résultat soit un roi sachant qu'on a tiré un pique est :
$$P_A(B)=\frac{P(A\cap B)}{P(A)}=\frac{1}{32}:\frac{1}{4}=\frac{1}{8}.$$

On peut retrouver intuitivement ce résultat. En effet, sachant que le résultat est un pique, on a une chance sur 8 d'obtenir le roi.

b) Un sac contient 50 boules, dont 20 boules rouges et 30 boules noires, où il est marqué soit "Gagné" ou soit "Perdu"
Sur 15 boules rouges, il est marqué Gagné.
Sur 9 boules noires, il est marqué Gagné.
On tire au hasard une boule dans le sac.
Soit *R* l'événement "On tire une boule rouge".
Soit *G* l'événement "On tire une boule marquée Gagné"
Donc $R\cap G$ est l'événement "On tire une boule rouge marquée Gagné".

Alors : $P(R)=\dfrac{20}{50}=\dfrac{2}{5}=0,4$ et $P(R\cap G)=\dfrac{15}{50}=\dfrac{3}{10}=0,3$.
Donc la probabilité qu'on tire une boule marquée Gagné sachant qu'elle est rouge est :
$$P_R(G)=\frac{P(R\cap G)}{P(R)}=\frac{0,3}{0,4}=\frac{3}{4}=0,75.$$

On peut retrouver intuitivement ce résultat. En effet, sachant que le résultat est une boule rouge, on a 15 chances sur 20 qu'il soit marqué Gagné.

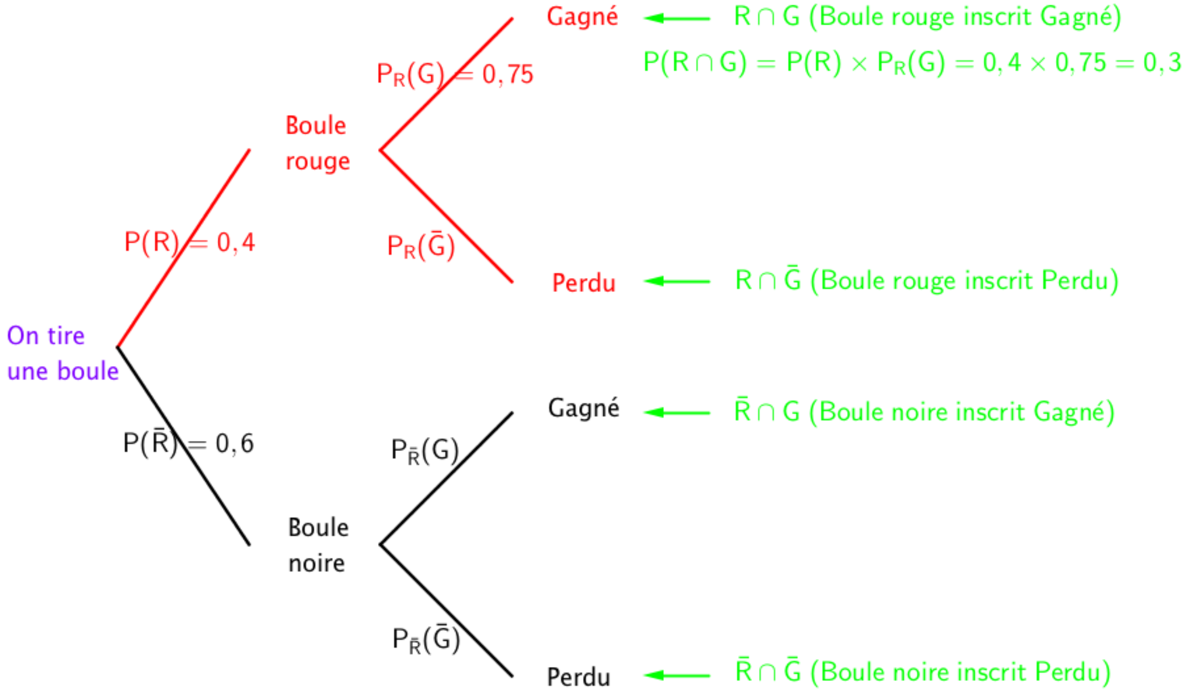
2. La probabilité conditionnelle suit les règles et lois de probabilités vues dans les classes antérieures. On a en particulier :

- $0\leqslant P_A(B)\leqslant 1$
- $P_A(\overline{B})=1-P_A(B)$
- $P(A\cap B)=P(A)\times P_A(B)$

II- Arbre pondéré :

1. Exemple :

On reprend le 2[e] exemple étudié au paragraphe I.
L'expérience aléatoire peut être schématisée par un arbre pondéré (ou arbre de probabilité) :



2. Règles :

- Règle 1 : La somme des probabilités des branches issues d'un même nœud est égale à 1.

Exemples :
- A partir du nœud "On tire une boule", on a : $P(R)+P(\bar{R})=0,4+0,6=1$
- A partir du nœud "Boule rouge", on a : $P_R(\bar{G})=1-P_R(G)=1-0,75=0,25$.

- Règle 2 : La probabilité d'une extrémité d'un chemin est égale au produit des probabilités du chemin aboutissant à cette extrémité.

Exemple :
On considère l'extrémité $R \cap G$.
On a : $P(R \cap G)=P(R)\times P_R(G)=0,4\times 0,75=0,3$

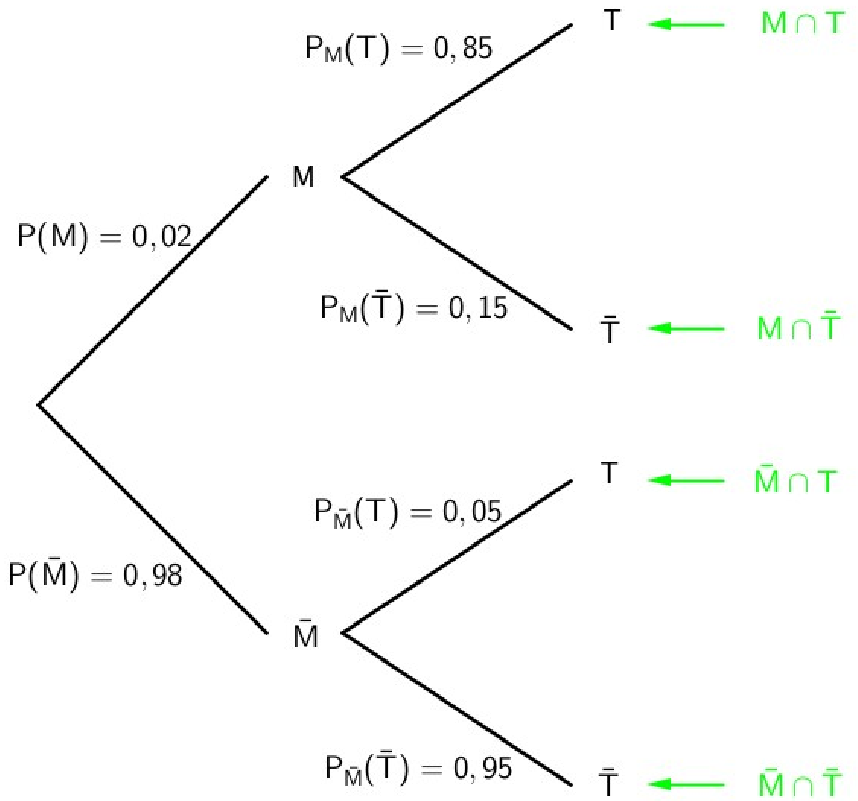
- Règle 3 (Formule des probabilités totales) : La probabilité d'un événement associé à plusieurs extrémités est égale à la somme des probabilités de chacune des ces extrémités.

Exemple :
L'événement "On tire une boule marquée Gagné" est associé aux extrémités $R\cap G$ et $\bar{R}\cap G$. On a :
$P(R\cap G)=0,3$ et $P(\bar{R}\cap G)=\frac{9}{50}=0,18$ (Probabilité de tirer une boule noire marquée Gagné)
Donc $P(G)=P(R\cap G)+P(\bar{R}\cap G)=0,3+0,18=0,48$.

Méthode : Calculer la probabilité d'un événement associé à plusieurs extrémités

Lors d'une épidémie chez des bovins, on s'est aperçu que si la maladie est diagnostiquée suffisamment tôt chez un animal, on peut le guérir ; sinon la maladie est mortelle.
Un test est mis au point et essayé sur un échantillon d'animaux dont 2 % est porteur de la maladie.
On obtient les résultats suivants :
– si un animal est porteur de la maladie, le test est positif dans 85 % des cas ;
– si un animal est sain, le test est négatif dans 95 % des cas.
On choisit de prendre ces fréquences observées comme probabilités pour toute la population et d'utiliser le test pour un dépistage préventif de la maladie.
On note respectivement *M* et *T* les événements « Être porteur de la maladie » et « Avoir un test positif ».

Un animal est choisi au hasard. Quelle est la probabilité que son test soit positif ?



La probabilité que le test soit positif est associée aux deux extrémités $M \cap T$ et $\bar{M} \cap T$.
$P(T) = P(M \cap T) + P(\bar{M} \cap T)$ (formule des probabilités totales)
= 0,02 x 0,85 + 0,98 x 0,05 = 0,066.
La probabilité que le test soit positif est égale à 6,6%.

III – Inversement du conditionnement :

Dans l'exercice précédent, la question intéressante est surtout de savoir si le test est fiable, c'est à dire connaître $P_T(M)$ et $P_{\bar{T}}(M)$. On utilise « à l'envers » les formules précédentes ou directement la formule de Bayes : $P_B(A) = \dfrac{P(A) \times P_A(B)}{P(B)}$ ce qui donne dans notre exemple :

$P_T(M) = \dfrac{P(M) \times P_M(T)}{P(T)} = \dfrac{0,02 \times 0,85}{0,066} = 0,26$ Seulement 25 % des tests positifs correspondent à des cas de maladie !

Cette formule est bien sur généralisable pour un plus grand nombre d'événements qui forment une partition de l'univers des possibles.

IV - Indépendance de deux événements :

1. On dit que deux événements *A* et *B* de probabilité non nulle sont indépendants lorsque $P(A \cap B) = P(A) \times P(B)$.

Remarque : On a également :
A et B sont indépendants, si et seulement si, $P_A(B) = P(B)$ ou $P_B(A) = P(A)$.

Exemple :
On tire une carte au hasard dans un jeu de 32 cartes.
Soit *R* l'événement "On tire un roi".
Soit *T* l'événement "On tire un trèfle".
Alors $R \cap T$ est l'événement "On tire le roi de trèfle".

On a :
$P(R) = \frac{4}{32} = \frac{1}{8}$, $P(T) = \frac{8}{32} = \frac{1}{4}$ et $P(R \cap T) = \frac{1}{32}$.
Donc $P(R) \times P(T) = \frac{1}{8} \times \frac{1}{4} = \frac{1}{32} = P(R \cap T)$.
Les événements *R* et *T* sont donc indépendants.
Ainsi, par exemple, $P_T(R) = P(R)$. Ce qui se traduit par la probabilité de tirer un roi parmi les trèfles et égale à la probabilité de tirer un roi parmi toutes les cartes.

Contre-exemple :
On reprend l'expérience précédente en ajoutant deux jokers au jeu de cartes.
Ainsi :
$P(R) = \frac{4}{34} = \frac{2}{17}$, $P(T) = \frac{8}{34} = \frac{4}{17}$ et $P(R \cap T) = \frac{1}{34}$.
Donc $P(R) \times P(T) = \frac{2}{17} \times \frac{4}{17} = \frac{8}{289} \neq P(R \cap T)$.
Les événements *R* et *T* ne sont donc pas indépendants.

Méthode : Utiliser l'indépendance de deux événements

Dans une population, un individu est atteint par la maladie *m* avec une probabilité égale à 0,005 et par la maladie *n* avec une probabilité égale à 0,01.
On choisit au hasard un individu de cette population.
Soit *M* l'événement "L'individu a la maladie *m*".
Soit *N* l'événement "L'individu a la maladie *n*".
On suppose que les événements *M* et *N* sont indépendants.
Calculer la probabilité de l'événement *E* "L'individu a au moins une des deux maladies".

$P(E) = P(M \cup N) = P(M) + P(N) - P(M \cap N)$
$= P(M) + P(N) - P(M) \times P(N)$, car les événements *M* et *N* sont indépendants.
= 0,005 + 0,01 – 0,005 x 0,01 = 0,01495

La probabilité qu'un individu choisi au hasard ait au moins une des deux maladies est égale à 0,01495.