

L'intelligence artificielle

Séance 5 : L'inférence bayésienne

Comment intervient l'inférence bayésienne dans l'apprentissage machine ou les diagnostics médicaux ?

Def : L'inférence bayésienne est une méthode de calcul de probabilités de causes à partir des probabilités de leurs effets

Un premier exemple d'inférence bayésienne :

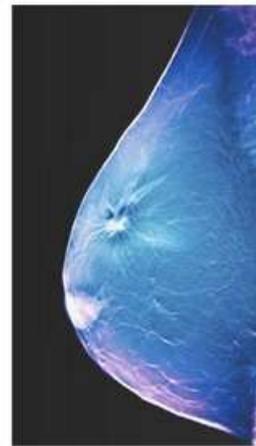
Pertinence d'un test médical

Sur 1 000 dossiers médicaux de femmes, un médecin trouve que 92 souffrent d'un cancer du sein. Un nouveau test d'imagerie médicale (a), géré par une **intelligence artificielle**, s'est révélé négatif pour 862 d'entre elles mais pourtant, parmi les patientes de ce groupe, 7 ont déclaré la maladie.

Le praticien aimerait savoir si ce nouveau test peut l'aider de façon fiable dans son diagnostic. Il établit alors le tableau à double entrée (b) de présentation des informations (dit **tableau de contingence**) afin de déterminer la fréquence d'individus testés positivement et réellement malades (individus « vrais positifs »).

... Pour mener une investigation

- Recopier et compléter le tableau ci-contre.
- Colorier de façon différente les cases correspondant aux vrais positifs (VP), faux positifs (FP), vrais négatifs (VN) et faux négatifs (FN).
- Calculer la fréquence des vrais positifs (VP/nombre total de positifs).
- Sachant qu'un test de détection sera considéré comme fiable si le taux de détection des vrais positifs est supérieur à 90 %, indiquer si le praticien peut se faire aider par cette IA.



a Détection de tumeurs par tomographie à rayons X (fausses couleurs).

Test	Malade	Sain	Total
Positif			
Négatif			
Total			

b Tableau de contingence.

Un autre exemple :

Le cas des maladies rares

La thalassémie est une maladie génétique qui affecte la production d'hémoglobine*. La fréquence d'apparition de l'une des formes de cette maladie est d'environ une personne sur 10 000.

Un laboratoire pharmaceutique a élaboré un test de dépistage et présente les résultats de tests effectués à partir des échantillons de sang de 1 000 patients (a). Le service d'études statistiques alerte la direction car le pourcentage de patients réellement malades quand le test est positif n'est que de 4,7 %.

On cherche à comprendre comment ce service a pu parvenir à ce résultat en utilisant une approche bayésienne.

... Pour mener une investigation

- Construire et compléter le tableau de contingence correspondant aux résultats du laboratoire en y exprimant les cas de VP, VN, FP et FN sous la forme de fréquence.

Test	Malade	Sain
Positif	990	2
Négatif	3	5

a Résultats du laboratoire.



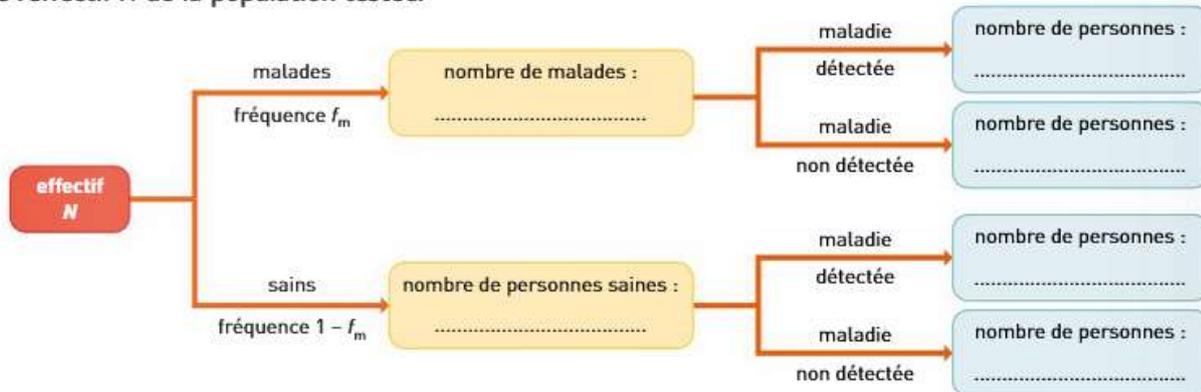
Repère

fréquence de non-apparition de la maladie = 1 – fréquence d'apparition de la maladie.

- Reproduire et compléter le tableau suivant en calculant les expressions de A , B et $A/(A + B)$.

A	fréquence d'apparition de la maladie \times fréquence (VP)	
B	fréquence de non-apparition de la maladie \times fréquence (FP)	
$A/(A + B)$	proportion de vrai malade quand le test est positif	

- Retrouver la formule $A/(A + B)$ en complétant le schéma suivant et montrer pourquoi elle est indépendante de l'effectif N de la population testée.



- Expliquer pourquoi le pourcentage de détection des VP de ce test est aussi faible dans la population.
- En suivant la même méthode, reprendre les calculs de détection des individus VP lorsque la fréquence d'apparition de la maladie est de $1/1\,000$ puis de $1/100$.
- Donner la relation qui existe entre fréquence d'apparition de la maladie et fiabilité du test.

POINT COURS :

► L'**inférence bayésienne** permet d'estimer la plausibilité d'une hypothèse, étant donné des observations concrètes. Elle est très utilisée en apprentissage automatique, notamment en vue de prononcer un diagnostic (médical, détection de spam...). Lorsqu'on applique une procédure de décision, nécessairement imparfaite, il apparaît des erreurs qu'on appelle des « faux », et qu'on qualifie de positif ou négatif selon le résultat de la détection. La construction d'un tableau de contingence permet de calculer la fréquence de faux positifs et de faux négatifs, qui caractérise la qualité de la procédure de décision.

Dans un test médical, si la fréquence de la maladie dans la population générale est f_m , alors le nombre de personnes réellement malades s'écrit (Fig. 5) :

$$\frac{f_m \cdot \text{fréquence(VP)}}{f_m \cdot \text{fréquence(VP)} + (1 - f_m) \cdot \text{fréquence(FP)}}$$

Test	Malade	Sain
Positif	vrai positif (VP)	faux positif (FP)
Négatif	faux négatif (FN)	vrai négatif (VN)

Fig. 5 : Exemple de tableau de contingence utilisé dans l'approche bayésienne.