Dépistage d'une maladie

1 Problématique

1. Imaginons que je suis testé négatif à un test de dépistage d'une maladie (Covid-19, SIDA, etc...), suis je **réellement** sain ?
   Ou dois-je réaliser un examen complémentaire ?
2. Et si je suis testé positif suis je **réellement** porteur de la maladie ? Ou dois-je là aussi réaliser un examen complémentaire ?

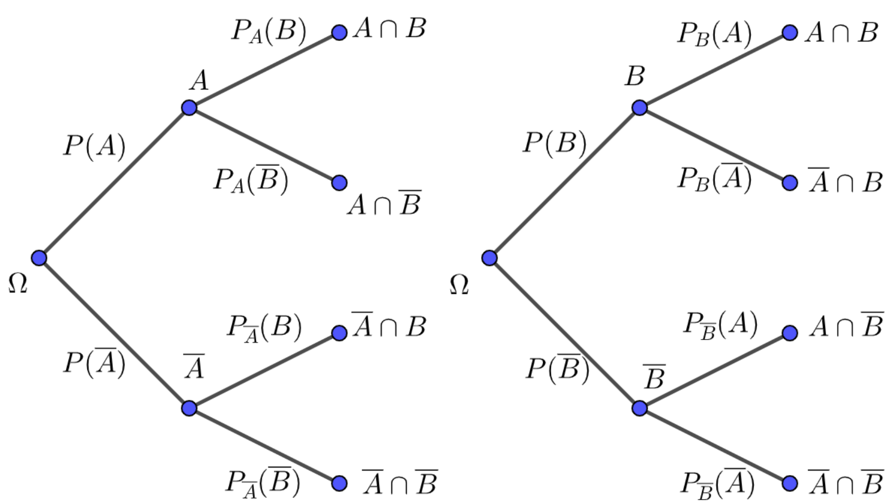
Pour répondre à ces questions nous allons utiliser le calcul des probabilités conditionnelles et le théorème de Bayes

2 Théorème de Bayes

Soit les évènements suivants :
— M :" le patient est atteint de la maladie M "
— T : " le patient est positif pour le test T servant à dépister la maladie M"

1. *Le patient est positif au test quelle est la probabilité qu'il soit malade ?*
   On cherche $P_T(M)$ appelée la valeur prédictive positive (VPP)
2. *Le patient est négatif au test quelle est la probabilité qu'il soit sain ?*
   On cherche $P_{\overline{T}}(\overline{M})$, appelée la valeur prédictive négative (VPN)

Pour trouver ces valeurs on se sert des arbres pondérés suivants



Théorème 1 $P_B(A) = \dfrac{P_A(B) \times P(A)}{P(B)} = \dfrac{P_A(B) \times P(A)}{P_A(B) \times P(A) + P_{\overline{A}}(B) \times P(\overline{A})}$

Preuve
$P(A \cap B) = P_A(B) \times P(A) = P_B(A) \times P(B) = \dfrac{P_A(B) \times P(A)}{P_A(B) \times P(A) + P_{\overline{A}}(B) \times P(\overline{A})}$

Donc $P_B(A) = \dfrac{P_A(B) \times P(A)}{P(B)} = \dfrac{P_A(B) \times P(A)}{P(A \cap B) + P(\overline{A} \cap B)} = \dfrac{P_A(B) \times P(A)}{P_A(B) \times P(A) + P_{\overline{A}}(B) \times P(\overline{A})}$

3 Sensibilité et spécificité d'un test

Lors de la phase de mise au point d'un test on procède

au près d'un échantillon de personnes que l'on sait porteurs de la maladie au dénombrement suivant :

1. VP le nombre de personnes malades ayant réagi positivement au test (appelé les vrais positifs)
2. FN le nombre de personnes malades ayant réagi négativement au test (appelé les faux négatifs)

puis on calcule **la sensibilité du test** $Se = \dfrac{VP}{VP + FN}$

au près d'un échantillon de personnes que l'on sait non porteurs de la maladie au dénombrement suivant :

1. FP le nombre de personnes saines ayant réagi positivement au test (appelé les faux positifs)
2. VN le nombre de personnes saines ayant réagi négativement au test (appelé les vrais négatifs)

puis on calcule **la spécificité du test** $Sp = \dfrac{VN}{VN + FP}$

Définition 1 *Dans le langage des probabilités*

*La **sensibilité** d'un test est* $Se = P_M(T)$

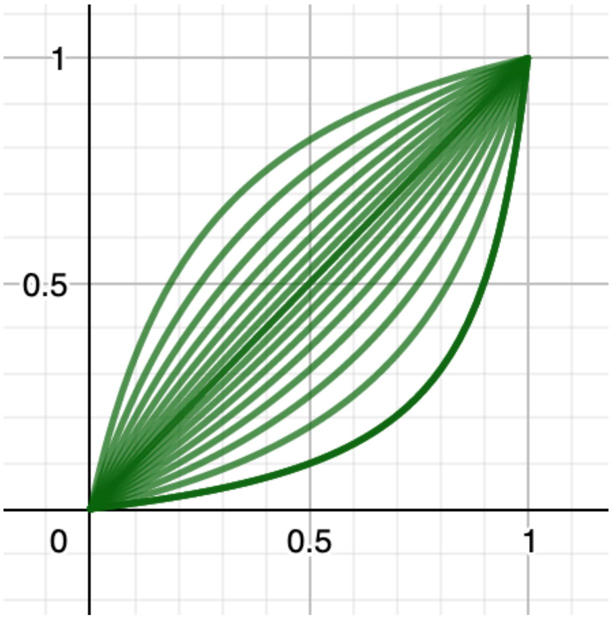
*La **spécificité** d'un test est* $Sp = P_{\overline{M}}(\overline{T})$

*La **prévalence** de la maladie est* $P(M)$

Théorème 2 *1.* $VPP = P_T(M) = \dfrac{Se \times P(M)}{Se \times P(M) + (1 - Sp) \times (1 - P(M))}$

2. $VPN = P_{\overline{T}}(\overline{M}) = \dfrac{Sp \times (1 - P(M))}{Sp \times (1 - P(M)) + (1 - Se) \times P(M)}$

4 Etude de fonctions



La valeur prédictive positive VPP et la valeur prédictive négative VPN sont des fonctions de la sensibilité et la spécificité d'un test et de la prévalence de la maladie

On a tracé ci-dessus la fonction f définie par $f(x) = \dfrac{sx}{(s+p-1)x+1-p}$ avec plusieurs valeurs de $s \in [0,1]$ la sensibilité, plusieurs valeurs de $p \in [0,1]$ la spécificité et $x \in [0,1]$ la prévalence

Mais une fois le test adopté la sensibilité et la spécificité ne varient plus ainsi on peut considérer VPP et VPN **uniquement fonction de la prévalence** avec comme **paramètres** la sensibilité et la spécificité

Quelle est la nature de cette fonction ? affine ? carrée ?

Théorème 3
1. *VPP et VPN sont des fonctions homographiques de la prévalence*
2. *La valeur prédictive positive VPP est une fonction croissante de la prévalence*
3. *La valeur prédictive négative VPN est une fonction décroissante de la prévalence*

Il existe un outil pour étudier les variations des fonctions "simples", la dérivation

5 Fonction dérivée

Théorème 4
1. *f est* ***croissante sur un intervalle I*** $\iff$ *la fonction dérivée de f notée f' est* ***positive*** *sur I*
2. *f est* ***décroissante sur un intervalle I*** $\iff$ *la fonction dérivée de f notée f' est* ***négative*** *sur I*

On rappelle en classe les principales techniques pour calculer la fonction dérivée d'une fonction "simple"

Théorème 5 *Une fonction homographique définie sur $\mathbb{R} - \{-\dfrac{d}{c}\}$ par $f(x) = \dfrac{ax+b}{cx+d}$ est croissante sur $]-\dfrac{d}{c};+\infty[$ si $ad-bc>0$*

Preuve
$f'(x) = \dfrac{a(cx+d) - c(ax+b)}{(cx+d)^2} = \dfrac{ad-bc}{(cx+d)^2}$
Application
$\text{VPP}(x) = \dfrac{sx}{sx+(1-p)(1-x)} = \dfrac{sx}{(s+p-1)x+1-p}$
On calcule le déterminant on trouve $s(1-p) > 0$ donc VPP est **croissante** sur [0,1]
$\text{VPN}(x) = \dfrac{p(1-x)}{p(1-x)+(1-s)x} = \dfrac{-px+p}{(1-s-p)x+p}$
On calcule le déterminant on trouve $p(s-1) < 0$ donc VPN est **décroissante** sur [0,1]

En conclusion

"**La sensibilité et la spécificité sont des indicateurs de référence pour déterminer si un test est de qualité (performance intrinsèque du test)**. Mais elles ne suffisent pas pour juger de la pertinence de l'utilisation d'un test diagnostic au sein d'une population donnée. En effet, cette pertinence est notamment dépendante de ce qu'on appelle la Valeur prédictive positive (VPP). Celle-ci varie selon la prévalence de la maladie : plus il y a de personnes malades dans la population testée, plus elle est élevée. **Autrement dit, plus un groupe a été exposé au virus, moins le risque d'erreur est grand. Et inversement, moins le groupe a été exposé, plus le risque de résultat erroné augmente**." (Covid 19, prérequis sur les tests, Haute autorité de santé - Mai 2020)

5.1 Exemples

* Cancer du sein :
`https://cancer-rose.fr/2016/11/13/cancer-du-sein-un-peu-de-technique/`
* Sida :
`https://www.who.int/hiv/mediacentre/news/hiv-misdiagnosis-qa/fr/index5.html`