

A propos des tests

Introduction

On se place dans le cadre d'un test binaire. Pour concrétiser l'exposé, on supposera qu'il s'agit d'un test biologique censé détecter la présence ou non d'une maladie. Ce test se fonde sur diverses mesures biologiques (x_1, x_2, \dots, x_k) et se conclut par le calcul d'une fonction discriminante dont le résultat s'appelle le score :

$$(x_1, x_2, \dots, x_k) \xrightarrow{f} [0 \ 1]$$

Sa valeur est alors comparée à un seuil S permet de classer la personne testée en positif ou négatif.

$$\begin{cases} f(x_1, x_2, \dots, x_k) \geq S \rightarrow + \\ f(x_1, x_2, \dots, x_k) < S \rightarrow - \end{cases}$$

Etude empirique sur une population témoin

Par population témoin, il faut comprendre un ensemble de personnes pour lesquelles on connaît :

1. Leur état : malade ou sain.
2. Leur réponse au test : positif ou négatif.

Matrice de confusion

Cette matrice indique les effectifs de la population testée croisée selon deux modalités : l'état réel (malade ou sain) et le résultat du test (+ ou —)

		Résultat du test		Σ
		+	—	
Etat	Malade	vp	fn	m
	Sain	fp	vn	s
Σ		p	n	t

- ✓ vp : nombre de vrais positifs.
- ✓ fp : nombre de faux positifs¹.
- ✓ fn : nombre de faux négatifs².
- ✓ vn : nombre de vrais négatifs.
- ✓ $m = vp + fn$: nombre de malades.
- ✓ $s = fp + vn$: nombre de personnes saines.

¹ En statistique, on parle alors d'erreur de 2^o espèce : acceptation de l'hypothèse « la personne est malade » alors qu'elle est fausse.

² En statistique, on parle alors d'erreur de 1^o espèce : rejet de l'hypothèse « la personne est malade » alors qu'elle est vraie.

- ✓ $p = vp + fp$: nombre de personnes testées positives.
- ✓ $n = fn + vn$: nombre de personnes testées négatives
- ✓ $t = m + s = p + n = vp + fn + fp + vn$: effectif total.

Une première idée pour mesurer la fiabilité du test est le taux d'erreur : $\frac{fp + fn}{t}$. Mais cet indicateur est trop grossier : il mélange les faux positifs et les faux négatifs, deux types d'erreurs dont les conséquences pratiques peuvent être très différentes...

Sensibilité, Spécificité

On définit alors :

1. La sensibilité du test comme le taux de vrais positifs parmi les malades :

$$T_{vp} = \frac{vp}{m}$$

Une sensibilité élevée indique qu'il existe peu de faux négatifs, autrement dit de malades non détectés.

2. La spécificité du test comme le taux de vrais négatifs parmi les personnes saines :

$$T_{vn} = \frac{vn}{s}$$

Une spécificité élevée indique qu'il existe peu de faux positifs, autrement dit de personnes saines détectés positives.

Il est clair qu'un bon test se doit d'avoir une sensibilité et une spécificité élevée. L'ennui est que, pour une fonction discriminante donnée, ces deux vertus n'évoluent pas dans le même sens. Si on augmente le seuil :

- ✓ Le nombre de personnes détectées positives va diminuer. On risque alors de voir des malades passer du statut de vrais positifs à celui de faux négatifs ce qui diminuera vp et par suite T_{vp} .
- ✓ Le nombre de personnes détectées négatives va augmenter. On risque alors de voir des personnes saines passer du statut de faux positifs à celui de vrais négatifs ce qui augmentera vn et par suite T_{vn} .

L'inverse se produit en cas de diminution du seuil. Et si on pousse à l'extrême :

- ✓ Avec un seuil de 0 tout le monde est déclaré positif. Ce qui implique $T_{vp} = 100\%$ et $T_{vn} = 0\%$.
- ✓ Avec un seuil de 1 tout le monde est déclaré négatif. Ce qui implique $T_{vp} = 0\%$ et $T_{vn} = 100\%$.

Exemple

Sur une population (fictive) de 100 personnes dont 25 malades on pratique le test. Les scores obtenus s'étalent de 0 à 0,99. Voici les résultats obtenus pour différents seuils³ :

✓ SEUIL 0,5

MATRICE DE CONFUSION			
	+	-	
Malade	25	0	25
Sain	29	46	75
	54	46	100
Sensibilité	100%		
Spécificité	61%		

C'est parfait pour la sensibilité : tous les malades sont détectés positifs. Mais, pour la spécificité c'est un désastre !

✓ SEUIL 0,65

MATRICE DE CONFUSION			
	+	-	
Malade	22	3	25
Sain	7	68	75
	29	71	100
Sensibilité	88%		
Spécificité	91%		

Comme annoncé en amont, l'augmentation du seuil provoque une forte amélioration de la sensibilité au prix d'une baisse de la spécificité.

✓ SEUIL 0,75

MATRICE DE CONFUSION			
	+	-	
Malade	17	8	25
Sain	3	72	75
	20	80	100
Sensibilité	68%		
Spécificité	96%		

Avec ce seuil on multiplie les faux négatifs pour aboutir à une sensibilité désastreuse !

³ Voir classeur Excel [courbeROC2.xlsx](#).

À titre indicatif, voici pour différents tests concernant la covid-19, les valeurs indiquées par le syndicat des biologistes⁴ :

Type de test	Sensibilité (Tvp)	Spécificité (Tvn)
RT-PCR	$\approx 100\%$	$\approx 100\%$
RT-LAMP salivaire	$\approx 74\%$	$\approx 100\%$
Antigénique	$\approx 70\%$	$\approx 96\%$
Sérologique	Entre 88% et 100% Selon la date du test	$\approx 100\%$

La Haute Autorité de Santé, pour valider un test exige une sensibilité $\geq 80\%$ et une spécificité $\geq 99\%$.

⁴ https://www.sdbio.eu/images/acymailing/SDB_tableau-repere-tests_covid-ok.pdf

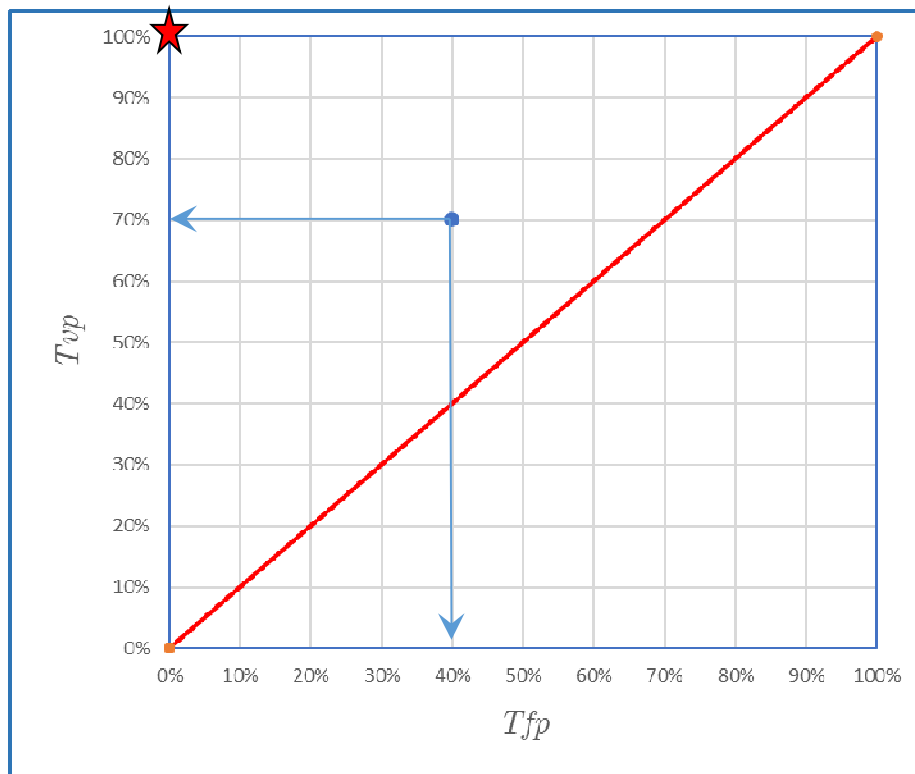
Courbe ROC⁵

L'influence du seuil sur la sensibilité et la spécificité suggère la démarche suivante :

1. Faire varier ce seuil.
2. Représenter par un point sur un graphique les valeurs obtenues pour ces deux critères.

En fait, pour des raisons de commodité, on a choisi, selon la valeur du seuil, de représenter :

1. En abscisse le taux de faux positifs $T_{fp} = 1 - T_{vn} = 1 - \text{Spécificité}$.
2. En ordonnée le taux de vrais positifs $T_{vp} = \text{Sensibilité}$.



On peut déjà mettre en exergue le point supérieur gauche du graphique de coordonnées (0%, 100%). Ce qui correspond à une sensibilité et une spécificité de 100% : le test idéal !

On a tracé aussi la diagonale. Ce qui correspond à $T_{vp} = T_{fp}$, autrement dit à un test totalement inutile.

⁵ ROC est l'acronyme de *Receiving Operating Characteristics* méthode issue des théories du signal.

Construction pratique

On s'appuie toujours sur notre population fictive de 100 personnes dont 25 malades. On connaît le score de chacun et on les range dans un tableau par ordre décroissant du score. Voici les premières lignes de ce tableau⁶ :

Il s'agit des valeurs pour le seuil égal au score sur la ligne.

Nb malades (1)		25					
NB sains (0)		75					
Individus	Etat	score	fp	vp	1-Spécificité	Sensibilité	Spécificité
					Tfp	Tvp	$1-Tfp$
θ^7		1,00	0	0	0%	0%	100%
1	1	0,99	0	1	0%	4%	100%
2	1	0,98	0	2	0%	8%	100%
3	1	0,97	0	3	0%	12%	100%
4	1	0,97	0	4	0%	16%	100%
5	1	0,97	0	5	0%	20%	100%
6	1	0,97	0	6	0%	24%	100%
7	1	0,96	0	7	0%	28%	100%
8	1	0,95	0	8	0%	32%	100%
9	1	0,92	0	9	0%	36%	100%
10	1	0,92	0	10	0%	40%	100%
11	1	0,90	0	11	0%	44%	100%
12	1	0,88	0	12	0%	48%	100%
13	0	0,85	1	12	1%	48%	99%
14	1	0,84	1	13	1%	52%	99%
15	1	0,82	1	14	1%	56%	99%
16	1	0,81	1	15	1%	60%	99%
17	0	0,81	2	15	3%	60%	97%
18	0	0,81	3	15	4%	60%	96%
19	1	0,79	3	16	4%	64%	96%
20	1	0,78	3	17	4%	68%	96%
21	1	0,74	3	18	4%	72%	96%
22	0	0,74	4	18	5%	72%	95%
23	1	0,73	4	19	5%	76%	95%
24	1	0,72	4	20	5%	80%	95%
25	0	0,71	5	20	7%	80%	93%
26	0	0,70	6	20	8%	80%	92%
27	1	0,68	6	21	8%	84%	92%
28	1	0,67	6	22	8%	88%	92%

Il reste à remplir les 5 colonnes de droite. Le principe est simple :

1. On part avec un seuil $S=1$ pour lequel il n'y a aucun positif (ni vrai ni faux) donc $fp = vp = 0$.
2. Ensuite on considère le seuil correspondant à l'individu 1 $S=0,99$. Il est le seul positif et est malade donc $fp = 0$ $vp = 1$.
3. Ensuite, on continue à descendre le seuil en suivant la colonne Score (entre 2 valeurs consécutives rien ne change). Pour $S=0,98$ $0,97 \dots 0,88$, on aura à chaque fois un positif de plus. Mais comme les 12 premiers sont malades, ce seront des vrais positifs dont le nombre s'incrémente tandis que les faux positifs restent à 0.

⁶ Pour le tableau complet voir classeur Excel [courbeROC2.xlsx](#).

⁷ L'individu θ est fictif.

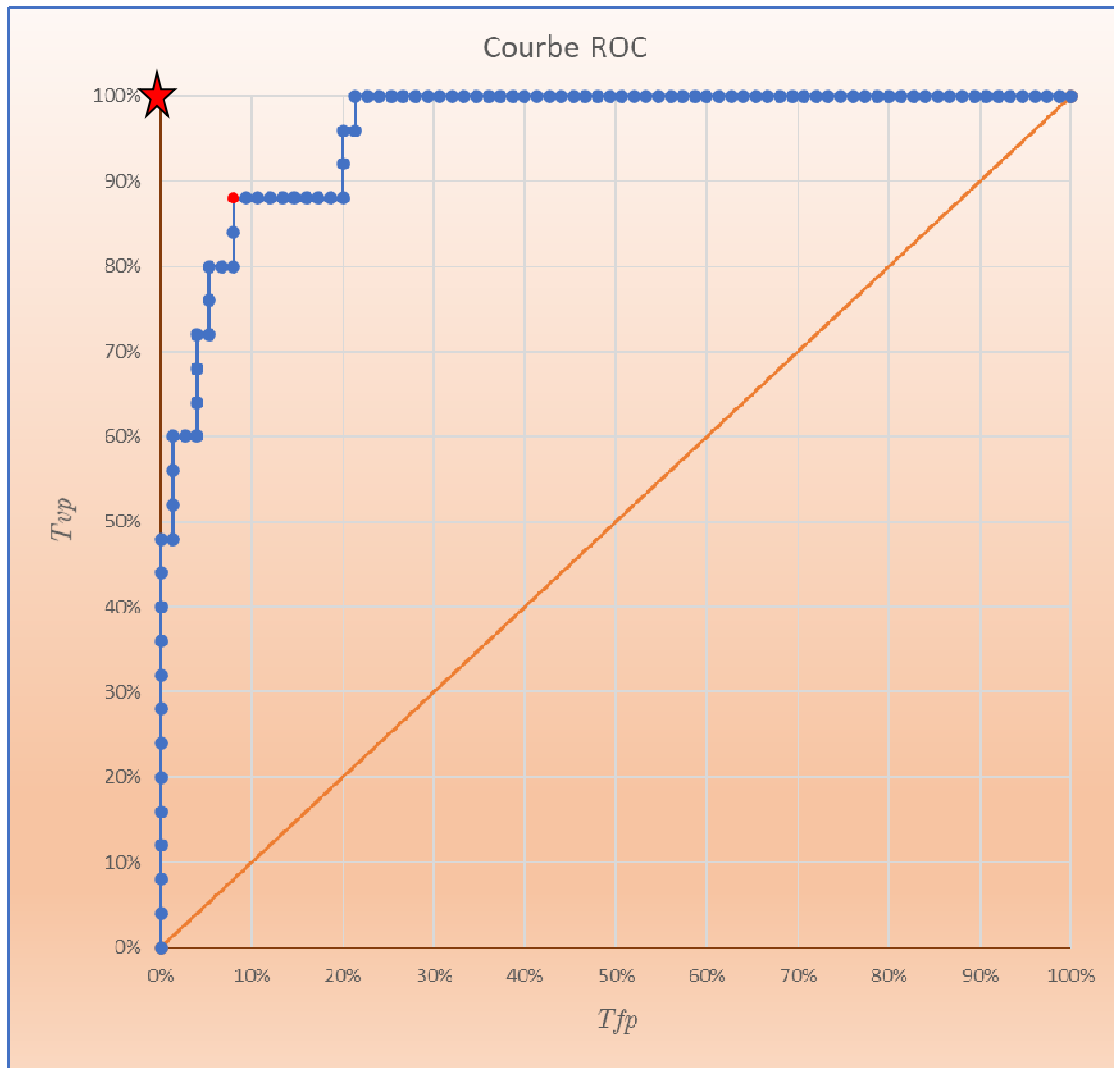
4. On arrive au seuil $S = 0,85$ correspondant à une personne saine, premier faux positif ce qui entraine $fp = 1$ $vp = 12$.

En résumé, sur la ligne de l'individu i , il suffit de dénombrer le nombre de 1 et le nombre de 0 dans la colonne État pour les individus $1 \dots i$. Il s'en suit que sur la dernière ligne on aura : $fp = s$ (nombre de personnes saines, 75 dans la population fictive) et $vp = m$ (nombre de malades, 25 dans la population fictive), ce qui donnera 100% pour les deux taux.

On peut présenter encore le procédé autrement. Pour chaque nouvelle ligne :

- ✓ Soit le nouvel individu est malade (1) et on incrémente vp .
- ✓ Soit il est sain (0) et on incrémente fp .

Ces calculs peuvent aisément mis en scène sous Excel en se servant, par exemple, de la fonction `NB.SI()`.



La courbe ROC de la population fictive

La courbe ROC se présente comme une fonction en escalier⁸ croissante de (0%, 0%) à (100%, 100%). Elle permet de répondre graphiquement à la question :

Quel seuil donne le meilleur compromis entre la sensibilité et la spécificité ?

On a souligné précédemment que le (0%, 100%) correspondait au test « idéal ». D'où l'idée de déterminer le point de la courbe le plus proche du coin supérieur gauche. Marqué en rouge sur le graphique, il correspond à l'individu 28 et au seuil de 0,67 et donne une sensibilité de 88% et une spécificité de 92%. Il s'en suit que plus la courbe est proche de ce point « magique » et donc éloignée de la diagonale meilleur est le test.

Présentation plus formelle

Soit $E = (E_1, \dots, E_i, \dots, E_t)$ où $E_i = \begin{cases} 1 & \text{si l'individu } i \text{ est malade} \\ 0 & \text{sinon} \end{cases}$ la liste des états des t individus

rangés selon l'ordre décroissant de leur score⁹ Sc_i .

On peut observer qu'au seuil $S = Sc_i$, tous les individus $i+1 \dots t$ sont en dessous du seuil donc classés – et non concernés par le dénombrement de positifs (vrais pu faux).

Alors :

- ✓ Le nombre vp_i de vrais positifs au seuil $S = Sc_i$ est égal au nombre 1 dans la sous-liste (E_1, \dots, E_i) . D'où :

$$vp_i = \sum_{k=1}^i E_k \text{ et } Tvp_i = \frac{vp_i}{m}$$

- ✓ Le nombre fp_i de faux positifs au seuil $S = Sc_i$ est égal au nombre 0 dans la sous-liste (E_1, \dots, E_i) . D'où :

$$fp_i = i - vp_i \text{ et } Tfp_i = \frac{fp_i}{s} = \frac{i - vp_i}{s}$$

Area Under Curve

Cet éloignement de la diagonale peut se mesurer en calculant l'aire sous la courbe (AUC). La surface sous la courbe étant la réunion de rectangle de base $Tfp_i - Tfp_{i-1}$ et de hauteur Tvp_i , cette aire se calcule par la formule suivante :

$$AUC = \sum_{i=1}^t (Tfp_i - Tfp_{i-1}) Tvp_i$$

Lorsque $E_i = 1$, $Tfp_i - Tfp_{i-1} = 0$. Lorsque $E_i = 0$, $Tfp_i - Tfp_{i-1} = \frac{1}{s}$. On peut donc encore écrire :

$$AUC = \frac{1}{s} \sum_{\substack{i=1 \\ E_i=0}}^t Tvp_i = \frac{1}{m \times s} \sum_{\substack{i=1 \\ E_i=0}}^t vp_i$$

Pour notre exemple, le calcul (voir classeur Excel) donne $AUC \approx 0,958$.

⁸ Le fait qu'elle soit en escalier est une conséquence du fait qu'à chaque nouvelle ligne on incrémente soit fp soit vp .

⁹ Par convention, en cas d'égalité de score entre un malade et un bien portant, on placera le malade devant.

Relation avec les inversions dans la liste E

Dans la liste E , une inversion se produit à chaque fois qu'un 0 précède un 1. Le nombre d'inversions se calcule en dénombrant pour chaque 0 les 1 le suivent et en additionnant. Ce qui se traduit par la formule :

$$\text{Nombre d'inversions } NI = \sum_{\substack{i=1 \\ E_i=0}}^t \text{Nombre de 1 dans } (E_{i+1}, \dots, E_t)$$

Sachant qu'il existe $m \times s$ couples de type $(1, 0)$ possibles, la fréquence des inversions est :

$$\text{Fréquence d'inversions } FI = \frac{1}{m \times s} \sum_{\substack{i=1 \\ E_i=0}}^t \text{Nombre de 1 dans } (E_{i+1}, \dots, E_t)$$

En se souvenant que vp_i est égal au nombre 1 dans la sous-liste (E_1, \dots, E_i) :

$$AUC + FI = \frac{1}{m \times s} \sum_{\substack{i=1 \\ E_i=0}}^t \text{Nombre de 1 dans } (E_1, \dots, E_i) + \text{Nombre de 1 dans } (E_{i+1}, \dots, E_t)$$

$$AUC + FI = \frac{1}{m \times s} \sum_{\substack{i=1 \\ E_i=0}}^t \underbrace{\text{Nombre de 1 dans } E}_{=m} = \frac{1}{s} \sum_{\substack{i=1 \\ E_i=0}}^t 1 = 1$$

Comme il existe s 0 dans E , $\sum_{\substack{i=1 \\ E_i=0}}^t 1 = s$. Et finalement :

$$\boxed{AUC = 1 - FI}$$

L'aire sous la courbe est d'autant plus grande que les inversions sont rares dans la liste E .

Les valeurs prédictives

À l'instar des définitions de sensibilité et spécificité, on définit aussi :

1. La Valeur prédictive positive, comme étant le taux de malades parmi les positifs

$$\boxed{Vpp = \frac{vp}{p}}$$

2. La Valeur prédictive négative, comme étant le taux de personnes saines parmi les négatifs :

$$\boxed{Vpn = \frac{vn}{n}}$$

On conçoit bien l'intérêt de ces deux notions lorsqu'on appliquera le test non sur une population témoin dont on connaît tout mais sur une population à explorer.

Devant un test positif, le praticien se posera la question : quelle est la probabilité que le sujet soit réellement malade ?

Devant un test négatif, le praticien se posera la question : quelle est la probabilité que le sujet soit réellement sain ?

Mais ces valeurs prédictives soulèvent une difficulté. Supposons que l'on multiplie par un coefficient α le nombre de malades tout en respectant la proportion positif/négatif. La nouvelle matrice confusion sera :

		Résultat du test		Σ
		+	—	
Etat	Malade	$\alpha \times vp$	$\alpha \times fn$	$\alpha \times m$
	Sain	fp	vn	s
Σ		p'	n'	t'

La sensibilité $\frac{\alpha \times vp}{\alpha \times m} = \frac{vp}{m}$ et la spécificité $\frac{vn}{s}$ seront inchangées. Mais il n'en sera pas de même pour les valeurs prédictives. À l'encontre de la sensibilité et la spécificité, ces dernières dépendent de la proportion de malades dans la population, ce que l'on nomme la prévalence :

$$Pr = \frac{m}{t}.$$

Donc la population témoin ne fournira pas en se fondant sur les valeurs empiriques de Vpp et Vpn une estimation des probabilités correspondantes dans une population pour laquelle on ignore la prévalence de la maladie.

Circonstance aggravante, l'évaluation du test est souvent effectuée, non pas sur une population témoin, mais indépendamment sur deux : une constituée exclusivement de malades, l'autre exclusivement de sujets sains. Pour lever cet obstacle, il va falloir se plonger un peu dans le calcul des probabilités.

Un peu de probabilités

Probabilités conditionnelles

Si A et B sont deux évènements (de probabilité non nulle), la probabilité conditionnelle de A sachant (ou si) B est définie par :

$$P[A/B] = \frac{P[A \cap B]}{P[B]}$$

De manière symétrique est définie la probabilité de B sachant A :

$$P[B/A] = \frac{P[A \cap B]}{P[A]}$$

Rappelons, au passage, que les 2 évènements A et B sont dits indépendants si et seulement si :

$$P[A] = P[A/B] \Leftrightarrow P[B] = P[B/A] \Leftrightarrow P[A \cap B] = P[A] \times P[B]$$

Un problème courant rencontré en probabilité et statistique est que l'on sait calculer (ou estimer) facilement $P[B/A]$, mais que l'on est plus intéressé par $P[A/B]$. On peut écrire :

$$P[A/B] = \frac{P[A \cap B]}{P[B]} = \frac{P[A \cap B]P[A]}{P[A]P[B]} = \frac{P[B/A]P[A]}{P[B]}$$

Et, ensuite sachant que $B = (B \cap A) \cup (B \cap \bar{A})$ et que $B \cap A$ et $B \cap \bar{A}$ sont disjoints, on a :

$$P[B] = P[B \cap A] + P[B \cap \bar{A}] = P[B/A]P[A] + P[B/\bar{A}]P[\bar{A}]$$

Et finalement, on obtient la formule des Bayes :

$$P[A/B] = \frac{P[B/A]P[A]}{P[B/A]P[A] + P[B/\bar{A}]P[\bar{A}]}$$

Application aux valeurs prédictives

On considère les 2 variables aléatoires : $M \in \{1, 0\}$ $T \in \{+, -\}$ représentant respectivement l'état de la personne et le résultat du test.

- ✓ $P[T = +/M = 1]$ correspond à la sensibilité Tvp .
- ✓ $P[T = -/M = 1]$ correspond au taux de faux négatif $Tfn = 1 - Tvp$
- ✓ $P[T = -/M = 0]$ correspond à la spécificité Tvn .
- ✓ $P[T = +/M = 0]$ correspond au taux de faux positifs $Tfp = 1 - Tvn$.
- ✓ $P[M = 1/T = +]$ correspond à la valeur prédictive positive Vpp .
- ✓ $P[M = 0/T = -]$ correspond à la valeur prédictive négative Vpn .
- ✓ $P[M = 1]$ correspond à la prévalence Pr .

Par application de la formule de Bayes, on a pour la valeur prédictive positive :

$$P[M = 1/T = +] = \frac{P[T = +/M = 1]P[M = 1]}{P[T = +/M = 1]P[M = 1] + P[T = +/M = 0]P[M = 0]}$$

En allégeant les notations :

$$Vpp = \frac{Tvp \times Pr}{Tvp \times Pr + (1 - Tvn) \times (1 - Pr)}$$

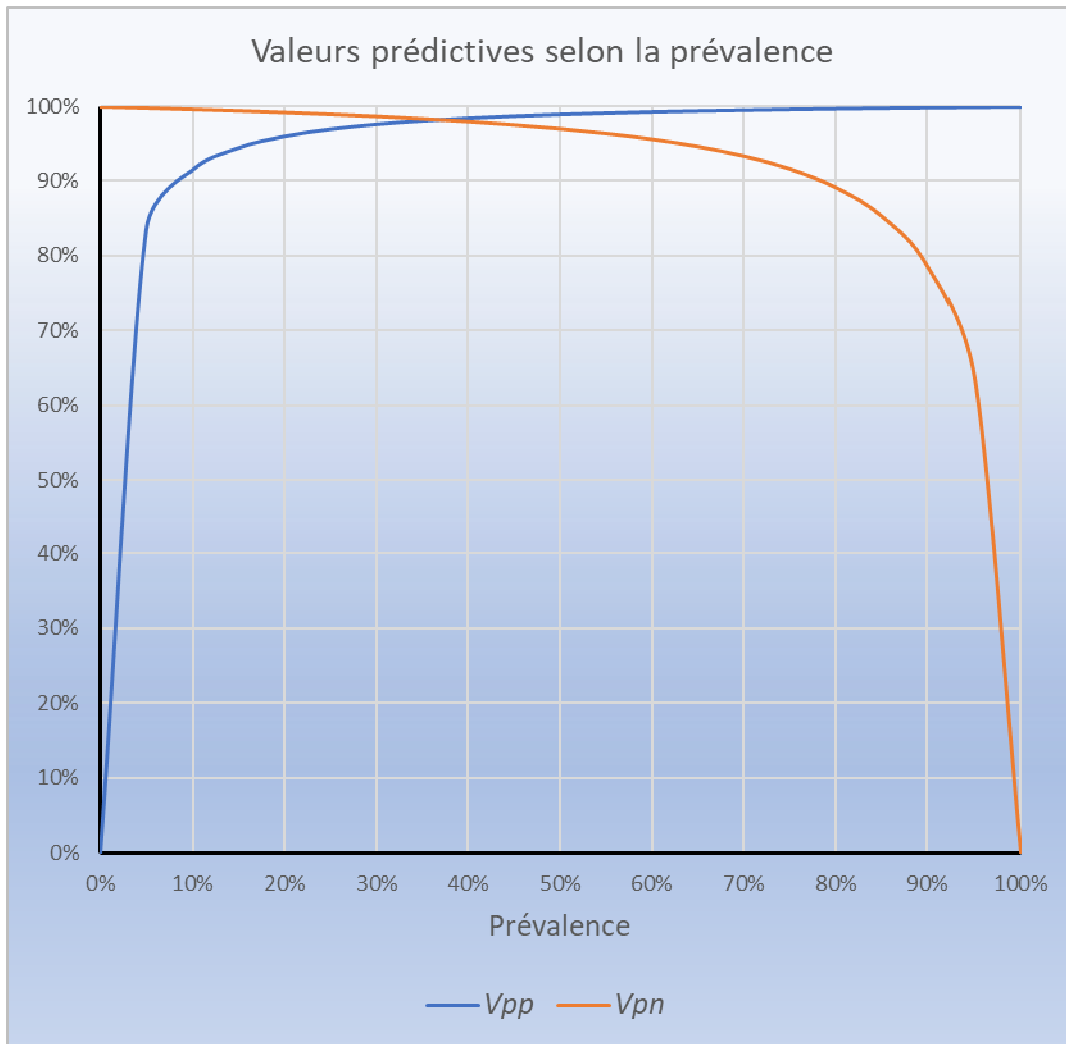
Par application de la formule de Bayes, on a pour la valeur prédictive négative :

$$P[M = 0/T = -] = \frac{P[T = -/M = 0]P[M = 0]}{P[T = -/M = 0]P[M = 0] + P[T = -/M = 1]P[M = 1]}$$

En allégeant les notations :

$$V_{pn} = \frac{T_{vn} \times (1 - Pr)}{T_{vn} \times (1 - Pr) + (1 - T_{vp}) \times Pr}$$

Voici pour une sensibilité de 97% et une spécificité de 99% les courbes représentatives des valeurs prédictives en fonction de la prévalence¹⁰ :



On notera que si la prévalence est nulle ($Pr = 0\%$), alors $V_{pp} = 0\%$, $V_{pn} = 100\%$. Ce qui est frappé du sceau du bon sens, puisqu'alors ; en absence de malades, la probabilité d'être atteint est nulle et celle de ne pas l'être de 100% quel que soit le résultat du test !

Remarque symétrique si la prévalence est de 100%.

On notera aussi que dans le cas d'un test « parfait » ($T_{vp} = T_{vn} = 100\%$), alors $V_{pp} = 100\%$, $V_{pn} = 100\%$.

¹⁰ Voir classeur Excel [courbeROC2.xlsx](#).

Quelques références

- Rakotomalala Ricco *Courbe ROC*
http://eric.univ-lyon2.fr/~ricco/cours/slides/roc_curve.pdf
- Rakotomalala Ricco *Quelques courbes pour évaluer les performances des classifieurs*
http://eric.univ-lyon2.fr/~ricco/tanagra/fichiers/fr_Tanagra_Spv_Learning_Curves.pdf
- Desquilbet Loïc *Tutoriel sur les courbes ROC et leur création grâce au site InterneteasyROC*
<https://hal.archives-ouvertes.fr/hal-02870055>
- Syndicat des Biologistes *Tableau-repère sur les tests diagnostiques covid 19*
https://www.sdbio.eu/images/acymailing/SDB_tableau-repere-tests_covid-ok.pdf
- Fourati S,
Audureau E,
Chevaliez S,
Pawlotsky JM *Évaluation de la performance diagnostique des tests rapides d'orientation diagnostique antigéniques COVID-19 AP-HP*