

Fiabilité des tests diagnostiques

Huguette Croisier

Version provisoire (novembre 2020)

Introduction

Vous avez peut-être vu passer sur le web des articles et/ou vidéos mettant en garde, avec des exemples assez frappants (Klein, 2020; Louape, 2012; Thibaut, 2016; Yungster, 2020), contre la confiance qu'on peut accorder "en pratique" à des tests fiables "en théorie" pour diagnostiquer la présence d'une maladie chez un individu¹. Je voudrais ici contribuer à mieux cerner la portée de ces arguments, en particulier dans le cas d'actualité brûlante du covid-19. Je me baserai pour ce faire sur l'excellent article du blog "Les Zétérinaires" à ce sujet (Chantreau, 2020). Je souhaite de plus expliciter, d'abord de manière intuitive puis mathématique, le raisonnement bayésien qui sous-tend les résultats donnés, de manière à ce que ce document puisse facilement être utilisé dans un cadre pédagogique comme exemple d'application du calcul des probabilités.

1 Concepts préalables

Tout d'abord, qu'entend-on par "fiabilité" d'un test diagnostique? En réalité, cette notion n'est pas quantifiable par un seul nombre, mais repose sur deux caractéristiques complémentaires d'un test diagnostique²:

- la *sensibilité*, c'est-à-dire la capacité du test à détecter "tous" les porteurs de la maladie. La sensibilité est liée à la notion de *faux négatif*: plus la sensibilité est grande, moins il y a de faux négatifs, c'est-à-dire de malades non détectés par le test.
- la *spécificité*, c'est-à-dire la capacité du test à ne détecter *que* les porteurs de cette maladie. La spécificité est liée à la notion de *faux positif*: plus la spécificité est grande, moins il y a de faux positifs, c'est-à-dire d'individus sains (ou porteurs d'une autre affection) identifiés erronément comme porteurs de la maladie.

¹Ce qui est entendu par "pratique" et "théorie" deviendra clair dans ce qui suit.

²Pour des informations sur la détermination de ces caractéristiques, voir par exemple l'excellent cours d'épidémiologie de T. Ancelle (Ancelle, 2014).

Dans le langage des probabilités, on peut résumer ces informations de la manière suivante:

$$\text{Sensibilité} = \mathbb{P}(\oplus|\textit{malade}) = 1 - \mathbb{P}(\ominus|\textit{malade}) = 1 - \mathbb{P}(\text{faux négatif})$$

$$\text{Spécificité} = \mathbb{P}(\ominus|\textit{sain}) = 1 - \mathbb{P}(\oplus|\textit{sain}) = 1 - \mathbb{P}(\text{faux positif})$$

où la notation $\mathbb{P}(A|B)$ représente la probabilité de l'événement A étant donné l'événement B (on parle de *probabilité conditionnelle*). Le symbole \oplus signifie que le résultat du test est positif, et le symbole \ominus signifie que le résultat du test est négatif.

2 Exemples

Considérons le cas d'un test diagnostique dont la sensibilité est de 70% et la spécificité de 99%, qui sont les ordres de grandeur de ces caractéristiques pour les tests RT-PCR utilisés pour le diagnostic du covid-19³. Autrement dit, la probabilité de faux négatif est de 30% et la probabilité de faux positif est de 1%.

Lorsqu'on confronte quelqu'un avec cette dernière information, et qu'on lui demande ce qu'il pense être sa probabilité d'être malade s'il est testé positif, l'immense majorité des individus répondent: 99%⁴. Cette réponse est incorrecte, et la probabilité réelle peut s'écarter drastiquement de cette valeur. On peut le comprendre assez facilement sur l'exemple suivant, typique des exemples "frappants" circulant sur le web.

Supposons que la prévalence de la maladie dans la population générale est de 1%, et que l'individu est testé dans le cadre d'un dépistage massif, c'est-à-dire sans suspicion particulière qu'il soit malade. Supposons également que la population en question compte 10000 individus (ce nombre n'a pas d'importance, il est choisi pour la commodité de calcul). Dans cette population, il y a donc $10000/100=100$ individus malades. Etant donné la sensibilité du test, (environ) 70 de ces 100 individus malades seront testés positifs et 30 seront testés négatifs. Avec une spécificité de 99%, on peut également estimer que parmi les 9900 individus non-malades, $0,99 \times 9900 = 9801$ individus seront testés négatifs et $0,01 \times 9900 = 99$ individus seront testés positifs.

³En réalité, la spécificité des test RT-PCR est proche de 100%(Chantreau, 2020), mais pour l'exemple, il est plus pédagogique de ne pas prendre cette valeur extrême. L'estimation de la sensibilité a quant à elle été revue à la hausse: elle serait proche de 88% (Jarrom et al., 2020).

⁴En partant de la définition de la sensibilité, qui est la probabilité d'être testé positif si on est malade, on pourrait également être tenté de répondre "70%", mais cela n'est pas correct non plus. Cela revient à confondre $\mathbb{P}(\oplus|\textit{malade})$ et $\mathbb{P}(\textit{malade}|\oplus)$, comme cela est explicité dans la section suivante.

Avec ces estimations, on peut calculer correctement la probabilité d’être malade lorsqu’on est testé positif (également appelée *valeur prédictive positive*, ou VPP). En effet, celle-ci est donnée par la proportion d’individus malades parmi les individus testés positifs⁵, autrement dit par le rapport entre le nombre d’individus malades et testés positifs (70) et le nombre total d’individus testés positifs (70+99):

$$\text{VPP} = \mathbb{P}(\text{malade}|\oplus) = \frac{\text{nombre de “malades et positifs”}}{\text{Nombre de “positifs”}}.$$

Ainsi, la VPP vaut $70/(70+99)=0,41$, càd 41% et non... 99%. Il y a donc 59% (et non 1%) de “faux positifs”⁶!

Les articles/vidéos “sensationnalistes” s’arrêtent souvent à ce résultat et en concluent qu’un tel test n’est pas fiable. Dans le cas d’une maladie contagieuse comme le covid-19, il conduirait en particulier à mettre en quarantaine beaucoup trop d’individus en réalité non porteurs de la maladie.

Cette conclusion est trop hâtive pour plusieurs raisons. Tout d’abord, il est tout aussi important d’estimer la probabilité d’être malade lorsqu’on est testé négatif. Le nombre d’individus malades et négatifs étant égal à 30, et le nombre total d’individus négatifs valant 30+9801, cette probabilité vaut $30/9831=0,003$. Ainsi, la proportion effective de “faux négatifs” est extrêmement faible: 0,3%. La proportion complémentaire de “vrais positifs”, ou *valeur prédictive négative* (VPN), vaut quant à elle 99,7%. Autrement dit, le risque de considérer comme sains des individus en réalité infectés - beaucoup plus grave que le risque de faux positif dans le cas d’une maladie contagieuse - est quasi nul.

Ensuite, et c’est là le plus important, ces résultats changent considérablement si la prévalence de la maladie dans la population est plus grande, ou si le test n’est pas effectué dans le cadre d’un dépistage massif. Par exemple, si on refait le raisonnement précédent en supposant une prévalence de 5%, on se rend compte que la VPP passe de 0,41 à 0,79 (elle devient donc supérieure à la sensibilité, qui vaut 0,7). Mais surtout, ce raisonnement ne tient pas compte du fait que la plupart des tests diagnostiques ne sont pas effectués “au hasard” (c’est le cas seulement des campagnes de dépistage massif), mais bien sur base d’une suspicion préalable de contamination (soit parce que l’individu présente des symptômes, soit parce qu’il a été en contact avec une personne contaminée). Or, le calcul effectué sur base de la prévalence n’est pas valable dans ce cas⁷. La manière la plus commode

⁵On peut notamment s’en convaincre au moyen de diagrammes de Venn [ref].

⁶La terminologie “faux positif” est généralement associée à la probabilité pré-test $\mathbb{P}(\oplus|\text{sain})$, mais nous l’employerons également pour référer à la probabilité post-test $\mathbb{P}(\text{sain}|\oplus)$, en distinguant les deux par les indices *a* (“ex-ante”) et *p* (“ex-post”) lorsqu’une confusion est possible.

⁷Ou, plus exactement, ce n’est plus la prévalence de la maladie dans la population générale qu’il faut considérer

de calculer les probabilités dans le cas général est introduite ci-dessous.

3 Formalisation du raisonnement

On montre assez facilement que les raisonnements effectués dans la section précédente sont équivalents mathématiquement à la formule de Bayes⁸:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Dans le cas des tests diagnostiques considéré ici, on a donc notamment:

$$\begin{aligned} P_p(FP) &= \mathbb{P}(sain|\oplus) = \frac{\mathbb{P}(\oplus \text{ et } sain)}{\mathbb{P}(\oplus)} = \frac{\mathbb{P}(\oplus|sain)\mathbb{P}(sain)}{\mathbb{P}(\oplus)} \\ P_p(FN) &= \mathbb{P}(malade|\ominus) = \frac{\mathbb{P}(\ominus \text{ et } malade)}{\mathbb{P}(\ominus)} = \frac{\mathbb{P}(\ominus|malade)\mathbb{P}(malade)}{\mathbb{P}(\ominus)} \end{aligned}$$

qui sont, respectivement, les probabilités d'être un faux positif (FP) et un faux négatif (FN) sur base du résultat du test (l'indice p signifie "post-test"). En utilisant les propriétés que

$$\begin{aligned} \mathbb{P}(\oplus) &= \mathbb{P}(\oplus|sain)\mathbb{P}(sain) + \mathbb{P}(\oplus|malade)\mathbb{P}(malade), \\ \mathbb{P}(\ominus) &= \mathbb{P}(\ominus|malade)\mathbb{P}(malade) + \mathbb{P}(\ominus|sain)\mathbb{P}(sain), \\ \mathbb{P}(sain) &= 1 - \mathbb{P}(malade) \end{aligned}$$

et l'abréviation $P_m = \mathbb{P}(malade)$, on peut encore écrire:

$$\begin{aligned} P_p(FP) &= \frac{(1 - Sp) (1 - P_m)}{(1 - Sp) (1 - P_m) + Se P_m} = 1 - VPP \\ P_p(FN) &= \frac{(1 - Se) P_m}{(1 - Se) P_m + Sp (1 - P_m)} = 1 - VPN \end{aligned}$$

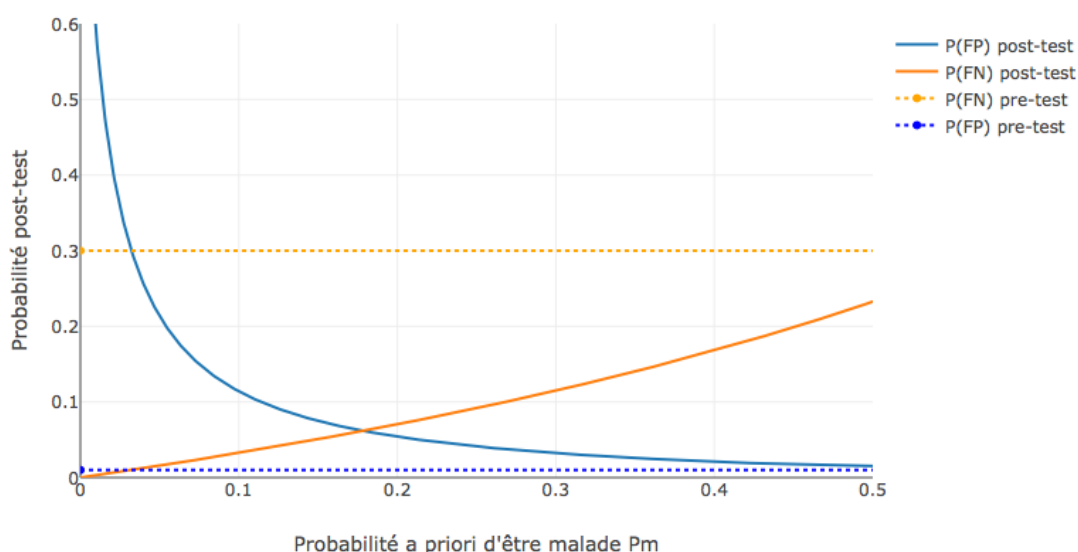
où Se est la sensibilité et Sp la spécificité du test telles que définies plus haut. On vérifie facilement que l'application de ces formules permet de retrouver les résultats de la section précédente (sans devoir spécifier la taille de la population), à condition de remplacer P_m par la prévalence de la maladie dans la population. Cependant, cette probabilité P_m n'est égale à la prévalence de la

dans ce cas, mais bien sa prévalence dans la population d'individus présentant des indices similaires de contamination.

⁸En effet, dans une population donnée, $\mathbb{P}(A|B) = \frac{\text{Nombre de "A et B"}}{\text{Nombre de "B"}} = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)}$ où la deuxième égalité est obtenue en divisant numérateur et dénominateur par la taille de la population. Par conséquent, on a aussi $\mathbb{P}(A \text{ et } B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$. Pour plus de détails, voir aussi ces excellentes vidéos: (Hoang, 2019; Sanderson, 2019).

maladie dans la population globale que dans le cas d'un dépistage massif, où l'individu est testé sans suspicion particulière qu'il soit infecté. De manière plus générale, P_m est la probabilité qu'a l'individu d'être malade avant d'être testé (on parle de probabilité pré-test, ou probabilité *a priori*). Lorsque l'individu est testé sur base d'une suspicion de contamination, parce qu'il présente des symptômes de maladie et/ou a été en contact avec des personnes malades, il faut substituer à la prévalence une *estimation* de la probabilité pré-test qu'il a été contaminé. Bien entendu, cette estimation n'est pas facile à réaliser - elle va dépendre des informations à disposition du médecin et de son expérience préalable avec la maladie (identification correcte des symptômes,...). C'est pourquoi on dit parfois que l'approche bayésienne du calcul des probabilités est "subjective". L'estimation faite dépend en tout cas crucialement des informations à disposition pour effectuer les calculs.

Pour fixer les idées, si on estime à 50% la probabilité pré-test P_m de l'individu d'être malade, avec les mêmes valeurs de sensibilité et de spécificité que dans la section précédente, la probabilité post-test de faux positif $P_p(FP)$ ne vaut plus que 1% (elle est donc égale à la probabilité pré-test), tandis que la probabilité de faux négatif $P_p(FN)$ vaut alors 23% (tandis que la probabilité pré-test est de 30%). Le graphique ci-dessous illustre cette dépendance drastique des probabilités post-test de faux positif et faux négatif à la probabilité *a priori* d'être malade P_m . Une version dynamique de ce graphique, permettant de visualiser l'influence d'un changement de sensibilité et/ou de spécificité sur l'allure des courbes, est disponible [ici](#).



4 Conclusions

- Une campagne de dépistage massif d'une maladie peu répandue dans une population n'a de sens que si on dispose de tests de sensibilité et de spécificité très grandes. Le même raisonnement s'applique à l'évaluation de la prévalence d'immunité au sein d'une population. En particulier, il est (pour l'instant) inutile d'effectuer des tests sérologiques à large échelle pour l'immunité vis-à-vis du covid-19 car la sensibilité et la spécificité des tests disponibles à cet effet est insuffisante (comparativement à la prévalence estimée pour l'immunité) (Chantreau, 2020).
- On peut améliorer considérablement la VPP et la VPN (càd diminuer les probabilités de faux positif et faux négatifs post-test) en combinant plusieurs tests. En effet, la probabilité P_m d'être malade avant le deuxième test est égale à la VPP du premier test lorsque celui-ci est positif, et est égale à la probabilité de faux négatif $1 - \text{VPN}$ dans le cas où le premier test est négatif (voir par exemple (Ancelle, 2014) pour une illustration). La VPP étant une fonction croissante de P_m et la VPN une fonction décroissante de P_m , l'application de deux tests en série résulte généralement en une amélioration de la fiabilité du diagnostic. C'est notamment pourquoi l'utilisation de tests antigéniques pour le diagnostic du covid-19, beaucoup plus rapides que les test RT-PCR mais de sensibilité moindre, est conditionnée à la réalisation d'un second test (RT-PCR) lorsque le résultat du premier test (antigénique) est négatif [ref].
- Si notre intuition est mise à mal par la formule de Bayes lorsque les probabilités *a priori* des événements sont faibles, on voit que cette constatation est nettement plus nuancée lorsque les probabilités a priori sont importantes (cf. graphique) - en particulier dans le cas d'une démarche diagnostique où les test sont effectués sur base d'un faisceau d'indications *préalables* que l'individu est porteur de la maladie. Il n'y a donc souvent pas lieu dans les faits de s'inquiéter de ces résultats contre-intuitifs, et de nombreux vulgarisateurs seraient plus avisés de pousser le raisonnement bayésien jusqu'au bout plutôt que d'alerter inutilement la population quant à la fiabilité des tests diagnostiques.

References

- Ancelle, T. (2014). Dépistage : performances d'un test. *Cours d'épidémiologie appliquée (youtube)*.
- Chantreau, J. (2020). Tests diagnostiques, comment! Pas fiables?! *Les Zétérinaires (blog)*.
- Hoang, L. N. (2019). Les lois des probabilités | Bayes 2. *Science4all (youtube)*.

- Jarrom, D., Elston, L., and Washington, J. (2020). Effectiveness of tests to detect the presence of SARS-CoV-2 virus, and antibodies to SARS-CoV-2, to inform COVID-19 diagnosis: a rapid systematic review. *BMJ Evidence-Based Medicine*, In press (Oct. 1).
- Klein, E. (2020). *Tracts de Crise (n25): "Je ne suis pas médecin, mais..."* . Gallimard.
- Louape, D. (2012). Les probabilités conditionnelles (Bayes level 1). *Science étonnante (blog)*.
- Sanderson, G. (2019). Bayes theorem. *3Blue1Brown (youtube)*.
- Thibaut, G. (2016). LA LOI DE BAYES - Argument frappant 3. *Monsieur Phi (youtube)*.
- Yungster, N. (2020). Where You Live Affects What Your COVID-19 Test Means. *Medium.com (blog)*.