

Comment rationaliser les décisions

JOHN SWETS • ROBYN DAWES • JOHN MONAHAN

Des outils mathématiques d'aide à la prise de décisions améliorent les diagnostics en médecine ou dans l'industrie.

Une opacité sur une mammographie est parfois le signe d'une tumeur. Doit-on l'opérer? Un comité qui étudie les dossiers de mise en liberté conditionnelle est alerté par le comportement violent d'un détenu : doit-on, malgré tout, le libérer? Dans un aéroport, un technicien observe une image anormale lors d'une inspection aux ultrasons d'une structure d'avion. Est-elle le signe d'une fissure inquiétante?

Souvent, les décisions sont difficiles à prendre, car les données sont incomplètes ou ambiguës. En médecine, en sécurité publique, dans l'industrie, dans les administrations, dans l'armée, au gouvernement, les enjeux sont parfois considérables, car des vies humaines sont en jeu.

Toutefois, les méthodes d'aide à la décision sont presque les mêmes, quels que soient les domaines. Des améliorations méthodologiques dans un domaine profitent donc généralement aux autres domaines. Ces méthodes, dont deux au moins sont validées, sont malheureusement méconnues ou inutilisées. L'une d'entre elles améliore la «pertinence» d'une décision : elle augmente la probabilité que la décision prise soit la bonne. L'autre améliore l'«utilité» d'une prise de décision ; elle évite les alertes inutiles. Ces méthodes sont statistiques et concernent donc un grand nombre d'individus ; au niveau individuel, elles influencent le choix du traitement thérapeutique, mais n'établissent pas une prédiction de l'évolution de la maladie d'une personne.

La décision n'impose pas toujours la mathématisation des problèmes. Dans certains domaines, telles la médecine

ou la météorologie, où de nombreux facteurs interviennent, les outils statistiques ne sont qu'une aide à la décision, des examens complémentaires ou des données nouvelles facilitant les décisions. En revanche, dans d'autres domaines, les analyses statistiques se sont souvent révélées plus pertinentes que les jugements subjectifs, même formulés par des professionnels expérimentés.

Nous examinerons ici les décisions limitées à deux éventualités. Une tumeur est-elle présente, oui ou non? Le délinquant peut-il être libéré sans risque de récidive, oui ou non? L'aile de l'avion est-elle défectueuse, oui ou non? De nombreux problèmes font apparaître des choix entre plus de deux possibilités, mais les décisions oui/non prédominent. Toutefois, on est généralement confronté à quatre solutions (et non deux) : les «vrais positifs» (une opacité se révèle être une tumeur) et les «vrais négatifs» (l'absence d'opacité correspond à une absence de tumeur), où la décision prise est sans ambiguïté ; les «faux positifs» (il n'y a pas de tumeur, malgré une opacité) et les «faux négatifs» (il y a une tumeur sans opacité), pour lesquelles la décision prise risque de ne pas être la bonne.

Les outils du métier

Si les tests diagnostiques donnaient les réponses attendues, les statistiques d'aide à la décision seraient évidemment inutiles. Toutefois, les résultats bruts des tests doivent généralement être interprétés. Par exemple, les ophtalmologistes mesurent la pression intraoculaire pour dépister les glau-

comes, des compressions du nerf optique et un accroissement de la dureté du globe, qui conduisent parfois à la cécité. Une pression très faible indique un œil sain, tandis qu'une pression élevée signale un glaucome... mais les valeurs intermédiaires?

Les informations statistiques éclairent partiellement la situation. Pour simplifier, admettons que la pression soit la seule mesure diagnostique disponible pour le dépistage du glaucome. Supposons qu'une pression inférieure à 10 (avec des unités arbitraires) corresponde toujours à un œil sain, que les pressions supérieures à 40 témoignent toujours d'un glaucome ; les mesures comprises entre 10 et 40 s'observent aussi bien sur des yeux sains que sur des yeux malades.

Pour élaborer des outils diagnostiques fiables, on doit d'abord identifier une grande population d'individus dont on mesure la pression intraoculaire, puis déterminer quels individus ont un glaucome au cours d'une période donnée. Pour chaque valeur de la pression, on détermine alors la probabilité qu'une personne ait un glaucome. Ces probabilités, et d'autres considérations présentées plus loin, définissent un point de partage rationnel, ou seuil diagnostique : les scores supérieurs ou égaux à ce seuil conduisent à un diagnostic positif (le patient souffre d'un glaucome), les scores inférieurs à un diagnostic négatif (le patient ne présente pas de glaucome).

Évidemment, un seul test diagnostique procure moins d'informations qu'une combinaison de tests. Pour accroître la pertinence d'un diagnostic, le décideur a intérêt à effectuer plusieurs examens qui fournissent

chacun une information en donnant un poids plus important aux mesures les plus prédictives de la pathologie étudiée. Les algorithmes mathématiques qui définissent les meilleurs tests et qui calculent le risque de glaucome d'après l'analyse des résultats combinés sont des méthodes statistiques prédictives.

Les données totalement objectives, telle la mesure de la pression, ne sont pas les seules utiles pour l'amélioration de la pertinence des méthodes statistiques prédictives ; des impressions subjectives peuvent également être quantifiées. On les objective en réalisant, par exemple, une liste de critères perceptibles (comme l'irrégularité d'un grain de beauté éventuellement malin), classés sur une échelle de 1 à 5.

On estime la pertinence globale des méthodes statistiques prédictives en étudiant les courbes ROC (de l'an-

glais *Receiver Operating Characteristic*, qui signifie «courbes des caractéristiques d'efficacité»). Ces courbes ont été utilisées pour la première fois au cours de la Seconde Guerre mondiale quand on a étudié la fiabilité des radars : on voulait quantifier leur capacité à distinguer les interférences aléatoires (le bruit de fond) du signal indiquant réellement la présence d'avions ennemis.

Les analyses des courbes ROC donnent une évaluation de la proportion d'individus atteints (et non atteints) en fonction d'une valeur seuil choisie : quelle proportion d'individus réellement atteints par la maladie étudiée seront dépistés (vrais positifs)? Et quel pourcentage d'individus sains seront par erreur considérés comme malades (faux positifs, ou «fausses alarmes»)?

Pour chaque seuil choisi, on détermine la proportion relative de vrais

positifs par rapport aux faux positifs. On obtient ainsi un graphe dont la courbure indique la pertinence : plus elle est forte, plus la prédiction est pertinente, car plus le rapport des vrais positifs sur les faux positifs est élevé.

Les vrais positifs et les faux positifs ne sont pas les seuls résultats possibles. Un diagnostic fondé sur des seuils donne aussi des vrais négatifs (des individus correctement estimés sains) et des faux négatifs (des individus malades non dépistés). Ces résultats complémentaires des précédents sont ignorés lors de la construction des courbes ROC. Par exemple, une proportion de vrais positifs égale à 80 pour cent correspond à une proportion de faux négatifs égale à 20 pour cent, ce qui représente une erreur considérable (un malade sur cinq n'est pas dépisté).

Comme les méthodes diagnostiques qui distinguent parfaitement les



1. LES QUESTIONS auxquelles un expert doit répondre par oui ou par non abondent, en médecine comme dans de nombreux autres domaines. Des méthodes statistiques confirmées, mais sous-utilisées, augmentent les chances de faire le bon choix.