Comment les mathématiques permettent-elles de modéliser l'association entre un facteur de risque et une maladie ?

Introduction

Mon ambition étant de devenir médecin, j'ai donc voulu comprendre comment les chercheurs en médecine faisaient pour mesurer l'association entre un facteur de risque et une maladie. Par exemple, les médecins assurent que consommer du tabac augmente le risque de développer un cancer du poumon. Ils sont en fait en mesure de l'affirmer grâce à un concept mathématiques précis: l'odds ratio.

La présentation que je vais vous faire aujourd'hui consiste d'abord à présenter l'odds en théorie puis l'appliquer à une situation concrète, précisément celle de l'exemple du cancer du poumon.

1) C'est quoi un odds ?

Un odds est le rapport entre la probabilité qu'un événement se produise et la probabilité qu'il ne se produise pas, soit en langage mathématique, le rapport entre le succès p et l'échec $1-p$.
On a ainsi la formule : $odds = \frac{p}{1-p}$

Nous sommes donc ici face à une épreuve de Bernoulli. Une épreuve de Bernoulli est une expérience aléatoire à deux issues contraires notées succès et échec, la probabilité du succès valant p et celle de l'échec $1-p$.
Toutes les expériences aléatoires ne sont pas nécessairement des épreuves de Bernoulli. Un match de Tennis est une épreuve de Bernoulli par exemple avec victoire du joueur 1 ou défaite du joueur 1 tandis qu'un match de football n'est pas une épreuve de Bernoulli car le match nul est possible, occasionnant alors trois issues.

Prenons un exemple simple pour se fixer les idées et calculer un odds.

Lorsque je lance un dé équilibré à 6 faces, chaque face présente la même chance d'apparaître. Nous sommes donc en situation d'équiprobabilité. J'ai ainsi autant de chance d'obtenir l'évènement D « obtenir un deux » que l'évènement T « obtenir un trois ». On a ainsi que $P(D) = \frac{1}{6} = P(T)$ en raison de l'équiprobabilité.
Je peux évaluer la valeur de l'odds pour l'évènement D par exemple.

$$odds = \frac{P(D)}{P(\bar{D})} = \frac{P(D)}{1-P(D)} = \frac{1/6}{5/6} = \frac{1}{5}$$

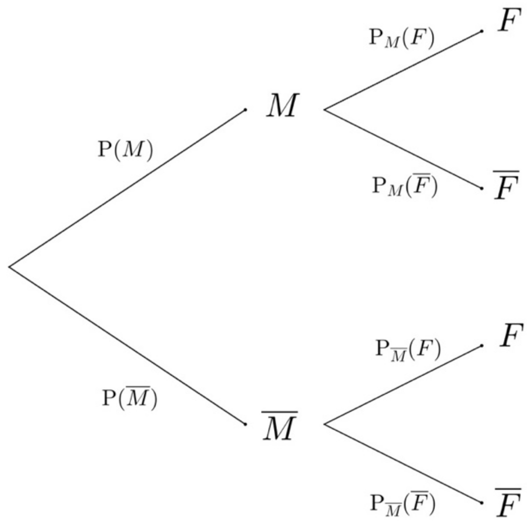
Traduction du résultat, cela veut dire qu'il y a 1 chance que j'obtienne le nombre 2 en lançant un dé équilibré contre 5 chances que je n'obtienne pas le nombre 2 que l'on peut traduire par "1 contre 5"
Ceci est d'ailleurs cohérent avec la traduction littérale du terme "odds" qui vient de l'anglais "chance".

2) Notion de probabilités conditionnelles

On revient désormais au problème posé qui consiste à mesurer l'association entre un facteur de risque et une maladie.
Dans une population donnée, on distingue 2 groupes de personnes:

- d'une part : le groupe malade qui correspond à l'événement M .
- d'autre part : le groupe sain qui correspond à l'événement contraire de M noté alors $\bar{M}$
- On note également F l'évènement être exposé au facteur de risque.

On peut ainsi construire un arbre de probabilité. On reconnait, les nœuds, les branches et les chemins, vocabulaire classique utilisé en probabilités. Les probabilités qui apparaissent sur les secondes branches représentent alors des probabilités conditionnelles. On note par exemple $P_M(F)$, la probabilité que le patient soit exposé au facteur de risque sachant qu'il est malade. On peut résumer ainsi la situation avec l'arbre complet que je vous ai reproduit sur mon document support.



Je me dois ici d'aborder le théorème de Thomas Bayes qui s'interprète comme une formule pour la probabilité conditionnelle d'une cause sachant qu'un évènement s'est produit. On a alors la formule des probabilités conditionnelles : $P_M(F) = \frac{P(F\cap M)}{P(M)}$

3) Calculs des odds

Dans le groupe des patients malades, l'odds est le rapport entre la probabilité d'être exposé au facteur de risque **sachant** que l'individu est malade sur la probabilité de ne pas être exposé au facteur de risque **sachant** que l'individu est malade. En conséquence, on a donc la formule :

$$odds\ (M) = \frac{P_M(F)}{P_M(\overline{F})}$$

On effectue le même calcul pour le groupe des patients en bonne santé :

$$odds\ (\overline{M}) = \frac{P_{\overline{M}}(F)}{P_{\overline{M}}(\overline{F})}$$

4) Calculs d'un odds ratio

Maintenant que nous avons compris ce qu'est un odds, le concept d'odds ratio s'éclaircit enfin. Ratio signifiant en latin rapport, un odds ratio est le rapport des odds calculés au sein de populations exposées diversement à un facteur de risque et plus précisément, ici, entre les 2 odds que nous venons de définir. En français, on traduit aujourd'hui odds ratio par rapport de cote. On peut donc calculer l'odds ratio de notre situation avec une formule qui peut inquiéter, formule que je vous ai reproduite sur mon document support.

$$OR = \frac{odds\ (M)}{odds\ (\overline{M})} = \frac{\frac{P_M(F)}{P_M(\overline{F})}}{\frac{P_{\overline{M}}(F)}{P_{\overline{M}}(\overline{F})}} = \frac{P_M(F) \times P_{\overline{M}}(\overline{F})}{P_M(\overline{F}) \times P_{\overline{M}}(F)}$$

L'odds ratio s'interprète souvent comme un risque relatif. Il est par exemple utilisé dans les paris sportifs. Ainsi si un cheval est côté à 3 contre 1, il a une chance sur 4 de gagner.

5) *Interprétation d'un odds ratio*

On peut donc interpréter un odds ratio en fonction de sa position par rapport à 1. Ainsi :

- Si $OR = 1$ il n'y a pas d'association entre le facteur de risque et la maladie. En d'autres termes, l'exposition au facteur de risque n'influence en rien le risque de développer la maladie. Il n'y a pas de corrélation.
- Si $OR > 1$ il y a association positive entre le facteur de risque et la maladie ; il y a donc augmentation du risque de développer la maladie dans le groupe exposé.
  Par exemple, si $OR = 2$, cela signifie qu'être exposé au facteur de risque va multiplier par 2 le risque de développer la maladie.
- Si $OR < 1$ il y a association négative entre le facteur de risque et la maladie ; on enregistre une diminution du risque de développer la maladie dans le groupe exposé : le facteur de risque est considéré comme "protecteur" contre la maladie Par exemple, si $OR = 0{,}5$ cela signifie qu'être exposé au facteur de risque va diviser par 2 le risque de développer la maladie

6) *Application à un cas concret*

Les données numériques utilisées ici sont issues d'une étude publiée en 2019 dans la revue scientifique ***The Lancet***, portant sur cette relation entre le cancer du poumon et la consommation de tabac. Cette étude a été menée sur deux populations comparables (âge, sexe, comorbidités) de même taille. Ainsi, les observations ont été réalisées sur 1000 hommes atteints du cancer du poumon et chez 1000 autres hommes du même âge, en bonne santé.
Il est important de rappeler que la notion de tabagisme n'est pas de fumer une cigarette mais de consommer régulièrement du tabac sous toutes ses formes.

↳ Ainsi parmi les patients malades :

- 90 % consomment du tabac, on a donc : $P_M(F) = 0{,}9$
- 10 % n'en consomment pas, on a donc : $P_M(\bar{F}) = 0{,}1$

$$odds\ (M) = \frac{P_M(F)}{P_M(\bar{F})} = \frac{0{,}9}{0{,}1} = \frac{9}{1}$$

Parmi les patients atteints du cancer du poumon, pour 9 personnes qui fument, 1 personne ne fume pas.

↳ Et parmi les patients sains :

- 60% consomment du tabac : ainsi, on a : $P_{\bar{M}}(F) = 0{,}6$
- 40% n'en consomment pas : et donc, on a : $P_{\bar{M}}(\bar{F}) = 0{,}4$

$$odds\ (\bar{M}) = \frac{P_{\bar{M}}(F)}{P_{\bar{M}}(\bar{F})} = \frac{0{,}6}{0{,}4} = \frac{3}{2}$$

Parmi les patients non malades, pour 3 personnes qui fument, 2 personnes ne fument pas. Cette deuxième affirmation est moins convaincante pour une campagne de sensibilisation par exemple.
On peut donc calculer l'odds ratio de notre situation à l'aide des résultats précédents.

$$OR = \frac{odds\ (M)}{odds(\bar{M})} = \frac{\frac{P_M(F)}{P_M(\bar{F})}}{\frac{P_{\bar{M}}(F)}{P_{\bar{M}}(\bar{F})}} = \frac{P_M(F) \times P_{\bar{M}}(\bar{F})}{P_M(\bar{F}) \times P_{\bar{M}}(F)} = \frac{0{,}9}{0{,}1} \times \frac{0{,}4}{0{,}6} = 6$$

Il apparaît ainsi que le risque de développer un cancer du poumon est 6 fois plus élevé chez les personnes qui consomment du tabac.

Pour traduire les résultats, je peux résumer ainsi la situation:
Si je suis une personne en bonne santé étant sur le point de fumer une cigarette, il est plus frappant :

- de savoir qu'en débutant un tabagisme, je multiplie mon risque de développer un cancer du poumon par 6.
- que de savoir que parmi les malades, 9 consommaient du tabac contre 1 qui n'en consommait pas. Je me sens moins concerné car je suis pour l'instant en bonne santé.

7) Mises en garde

L'odds ratio s'accompagne d'un intervalle de confiance. En effet, le chiffre odds ratio calculé ne peut pas être considéré comme étant tout à fait certain. Il existe alors une marge d'incertitude qui est précisée grâce à l'intervalle de confiance.
La notion d'intervalle de confiance est étudiée en classe de terminale via la loi binomiale. Lorsqu'on extrait une fréquence d'un échantillon, on peut extrapoler cette fréquence à une population globale à 95 % en encadrant la fréquence par deux bornes qui dépendent de la taille de l'échantillon. Naturellement, plus n est grand, plus l'amplitude de l'intervalle de confiance est petite.
Attention cependant à certains dangers !

- La proportion dans la population appartient à l'intervalle de confiance à 95 % sans pour autant pouvoir affirmer sa position dans cet intervalle.
- Si l'intervalle de confiance contient la valeur 1, alors le résultat n'est pas statistiquement significatif…

Rappelons ensuite que la recherche de facteur de risque est compliquée par la présence d'autres facteurs de risque. Dans le cas d'étude ici, on peut invoquer également la pollution de l'air ou l'exposition à l'amiante.

Enfin, il est important de préciser que pour les maladies où les facteurs de risques sont évidents, le calcul de l'odds ratio est important pour connaitre le réel impact du facteur de risque. Cependant, pour rechercher la ou les causes d'une maladie, c'est beaucoup plus compliqué que cela et l'IA va permettre des avancées importantes en brassant des quantités de données importantes.

Conclusion

En conclusion, les mathématiques servent grandement la médecine grâce à ce calcul d'odds ratio. Il est en effet essentiel d'orienter la prévention sur les facteurs de risque des maladies graves et de délivrer des messages clairs aux populations quant aux risques qu'elles encourent lorsqu'elles sont exposées à tel ou tel facteur de risque.