Société de Calcul Mathématique SA
Outils d'aide à la décision
depuis 1995

Méthodes probabilistes pour l'épidémiologie

1. Présentation du besoin

Peut-on avoir une approche probabiliste sur une question touchant à la santé ? ce n'est pas une évidence. Claude Bernard, dans son "Introduction à l'étude de la médecine expérimentale", 1865, est très critique à cet égard : "*En un mot, en se fondant sur la statistique, la médecine ne pourrait être jamais qu'une science conjecturale ; c'est seulement en se fondant sur le déterminisme expérimental qu'elle deviendra une science vraie, c'est-à-dire une science certaine.*"

Claude Bernard reconnaissait cependant que les probabilités pouvaient donner des renseignements précieux pour une population tout entière, ce qu'on appelle aujourd'hui épidémiologie, mais il considérait que leur apport, pour un individu donné, était faible. C'est certainement exact : nous n'avons en aucune manière l'ambition de nous substituer à l'art du médecin.

2. L'apport des méthodes probabilistes

Notre expérience est que les méthodes probabilistes ont un apport significatif dans les situations suivantes :

- Déceler l'existence d'un risque, pour une population donnée. Par exemple, à la demande de RTE, nous avons répondu à la question : y a-t-il un excès de leucémies de l'enfant, de maladies d'Alzheimer, au voisinage des lignes à haute tension ? (voir plus bas la liste de nos références) ;

- Apporter une information grossière, mais pertinente, à partir de données pauvres, insuffisantes et incomplètes : voir plus bas l'exemple traité à la demande de Peptinov ; cela vaut mieux qu'un résultats fin, mais et artificiel et erroné ;

- Mettre en garde les praticiens contre des erreurs de raisonnement, résultant de l'usage à l'aveugle de logiciels que l'on ne contrôle pas, en particulier de l'application erronée d'une hypothèse d'indépendance ;

- Permettre une comparaison avec ce que fait la Nature. Comme disait Claude Bernard : votre traitement guérit le malade, soit, mais la Nature aussi guérit le malade ; comment savoir si votre traitement fait mieux que la Nature ?

3. Le recueil des données

Lors d'essais cliniques, les panels disponibles ont généralement des effectifs faibles et toutes les configurations souhaitables (âge, poids, antécédents, etc.) ne sont pas représentées. Les données recueillies sont souvent imprécises et leur fiabilité n'est pas assurée. Le rôle que le hasard doit jouer : "randomisation" dans la détermination des sous-groupes et comparaison, du point de vue de l'efficacité, entre ce que fait la Nature et ce que fait le remède, est souvent peu claire. L'épidémiologie est un besoin essentiel de la société civile, mais les lois fondamentales des probabilités sont généralement mal connues des spécialistes de la discipline.

On a souvent recours à des méthodes de "randomisation", pour déterminer les groupes à qui on donnera un médicament et les groupes témoins.

Mais des erreurs conceptuelles majeures, intervenant dans la définition même des études épidémiologiques randomisées, sont souvent commises. Prenons un exemple simple pour faire comprendre ceci : un groupe constitué de 10 000 personnes, à égalité H/F. Vous extrayez un sous-groupe au hasard, mettons de taille 100 : la parité ne sera pas respectée ; la différence |H-F| sera en général de l'ordre de 10. De même, les profils des sous-groupes ne seront pas identiques, ni pour l'âge, ni pour quelque critère que ce soit. Le détail des erreurs commises, avec l'analyse mathématique correspondante, est donné dans notre article :
https://www.scmsa.eu/archives/BB_erreurs_etudes_randomisees_2020_09.pdf

Autre exemple de même nature : deux villes, en tous points identiques, même population ; admettons que la probabilité de cancer soit la même pour les deux ; l'application brutale des lois de probabilités usuelles indique que le nombre de cancers sera différent d'une ville à l'autre. Si cette différence naturelle n'est pas prise en compte, on aura tort d'en déduire que l'une des villes est plus exposée que l'autre, du fait de causes externes (lignes à haute tension, autoroutes, etc.).

4. L'exploitation des résultats

On donne un médicament à un certain groupe et un placebo à un autre groupe. On observe tant de guérisons dans le premier et tant de guérisons dans le second. Comment savoir si le médicament est efficace ? Les tests statistiques les plus communément utilisés reposent sur la "p-value", qui elle-même fait des hypothèses sur la loi de probabilité du phénomène en question : loi de probabilité qui est inconnue par définition. La p-value est maintenant critiquée par les scientifiques eux-mêmes : https://fr.wikipedia.org/wiki/Valeur_p.

De manière générale, il faut se garder d'avoir recours à des tests statistiques tout faits : tous reposent sur des hypothèses a priori, dont absolument rien ne garantit la légitimité.

5. La question de l'indépendance

Les outils statistiques usuels font une hypothèse d'indépendance entre événements. Les praticiens interprètent cette indépendance sous la forme : ce qu'on donne au patient A n'affectera

pas la réaction du patient B. Cette interprétation est erronée ; en probabilités, "indépendance" signifie qu'une information sur A ne donne aucune information sur B : ce n'est pas du tout la même chose.

Par exemple, il est communément admis que les activités humaines ont des cycles : suivant l'heure à laquelle vous aurez donné un médicament, il sera plus ou moins efficace. Supposons que vous ayez donné le médicament à tout un panel à la même heure : ce choix affectera tout le panel ; il n'y a pas indépendance.

Les erreurs de raisonnement de ce type peuvent avoir des conséquences très graves, conduisant à des erreurs de diagnostic. Une telle démonstration sera refusée par les autorités chargées de la validation.

6. Responsabilité juridique

Les recommandations que nous faisons ont une valeur juridique et nous en prenons la responsabilité. Même si elles reposent sur des méthodes probabilistes, elles seront utilisées de manière déterministe par le donneur d'ordre : je prescris ou je ne prescris pas tel ou tel médicament. L'expert accepte la présence d'incertitudes dans ses conclusions et il en rend compte à la justice. Voir à ce propos l'exposé que nous avons fait devant le Pôle Judiciaire de la Gendarmerie Nationale : https://www.scmsa.eu/archives/BB_presentation_expose_PJGN.pdf

Si, à l'inverse, les données ont été traitées par un logiciel "à l'aveugle", ou par une Intelligence Artificielle, aucune justification juridique n'est possible.

7. Nos références

- **Livres**

[IEPE] Bernard Beauzamy : Introduction à l'étude des Probabilités Expérimentales
https://www.scmsa.eu/livres/SCM_IEPE_order.htm

[MPPR] Bernard Beauzamy : Méthodes Probabilistes pour l'étude des phénomènes réels.
https://scmsa.eu/livres/SCM_MPPR_order.htm

- **Contrats traités**

En épidémiologie

2007-2008, Commissariat à l'Energie Atomique : Analyse critique des modèles en épidémiologie, relatifs aux effets des faibles doses de radioactivité ;
2009, Evaluation critique du rapport Bioinitiative ;
2009, Réseau de Transport d'Electricité : Analyse critique des modèles en épidémiologie, relatifs à la proximité des Lignes à Haute Tension ;
2010, EDF, Service des Etudes Médicales : Analyse des modèles de propagation pour cer-taines maladies, susceptibles de se développer au voisinage des centrales ;
2012, Aéroports de Paris : Les nuisances sonores autour des Aéroports de Paris et le processus politique d'évaluation ;

2012, ISTE, Nitrates et algues en Bretagne : Y a-t-il une relation ?
(travail présenté devant le 19ème Congrès National de la Coordination Rurale, 28/11/2012)
2013, Réseau de Transport d'Electricité : Analyse critique de l'étude "GEOCAP" (lien entre la leucémie de l'enfant et la présence de lignes HT) ;
2016, L'Oréal : Etude des données disponibles pour les accidents de la route entre le domicile et le lieu de travail ;
2017, Lien entre concentration d'ozone dans l'air et nombre d'admissions des consultations en kinésithérapie respiratoire, par Aurore Trébuchet, Alexis Cousin, Astrid Essartel, Adrien Schmitt et Bernard Beauzamy, Société de Calcul Mathématique SA ;
2020, Bernard Beauzamy : Analyse critique de l'article Stochasticité et épidémie de COVID-19 Groupe de modélisation de l'équipe ETE (Laboratoire MIVEGEC, CNRS, IRD, Université de Montpellier) ;
2024, Société Desialis : Dangerosité de l'Anthraquinone dans l'alimentation du bétail. Synthèse des connaissances disponibles et recommandations.

Compréhension des mécanismes physico-chimiques et de l'action sur le métabolisme

2011, Impeto Medical : Analyses statistiques relatives à la fonction sudorale
2015, RCIT : analyses statistiques en appui à une publication en cardiologie
2023, Peptinov : analyse probabiliste de données phase 1, immunothérapie active contre les maladies inflammatoires chroniques. Voir l'article : Méthodes mathématiques pour le traitement de données médicales imprécises :
https://www.scmsa.eu/archives/SCM_donnees_medicales_imprecises_2023.pdf