

Société de Calcul Mathématique SA

Outils d'aide à la décision

depuis 1995



L'enseignement des Probabilités

par Bernard Beauzamy

Professeur titulaire des Universités (1979-1995)

PDG, Société de Calcul Mathématique SA (depuis 1995)

email : contact@scmsa.com

Août 2016

Que ce soit au Lycée, où il commence à pénétrer, ou bien à l'Université et dans les Grandes Ecoles, où il est plus ancien, l'enseignement des probabilités est présenté de manière très académique. On commence, comme toujours, par une description des axiomes de base et on s'efforce de bâtir une théorie, mais, comme toujours, les axiomes et la théorie ne cadrent pas avec la réalité. Prenons une comparaison pour faire comprendre cette distance : la médecine s'appuie sur la biologie, et a besoin de comprendre le fonctionnement d'une cellule. Pourtant, il est impossible de définir axiomatiquement ce qu'est une cellule.

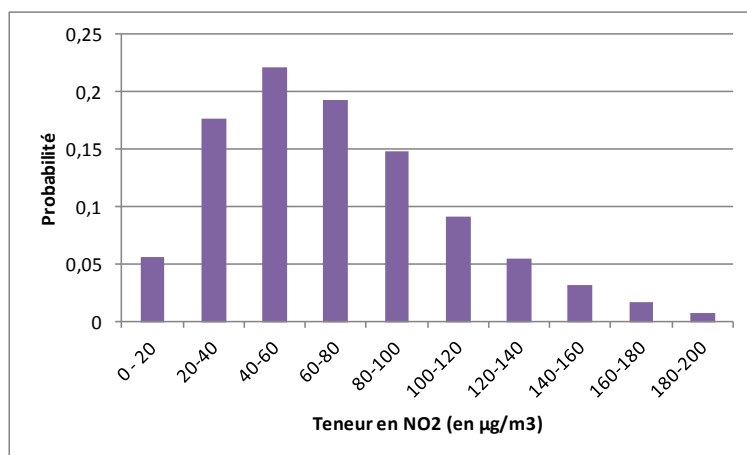
La première difficulté que l'on rencontre tient à la définition donnée dans l'enseignement : "une expérience dont le résultat dépend du hasard". Mais, déjà ici, l'utilisateur est perplexe : mais alors, ce que j'apprends ne s'appliquera qu'aux dés, aux cartes et aux urnes, parce que dans la vie courante rien ne dépend jamais du hasard !

Je vais illustrer mon propos par un exemple récent tiré d'un travail que nous avons réalisé pour le Ministère de l'Environnement, Direction Générale Energie Climat, Bureau Qualité de l'Air (2015) : il s'agissait d'étudier, pour le Boulevard Périphérique autour de Paris, les liens entre trafic (nombre de véhicules par heure) et pollution (présence mesurée de divers polluants émis par les véhicules, en particulier le monoxyde d'azote NO et le dioxyde d'azote NO₂) ; voir [1].

Chacun s'accordera à reconnaître que, dans une telle situation, le hasard n'intervient pas. L'expérience consiste en un certain nombre de mesures de trafic et de pollution, plus éventuellement des informations annexes (le vent, la pluviométrie, etc.).

Intéressons-nous simplement à la variable "pollution" ; sous la forme la plus simple, nous aurons un relevé, mettons horaire, de la concentration en NO₂ sur une certaine période. Cela se présente sous la forme d'un tableau à deux colonnes : première colonne la date, seconde colonne la concentration à cette date.

A partir de là, on réalise la "loi de probabilité" de la concentration en NO₂ : on divise les concentrations possibles en "classes", par exemple de 0 à 20 micro-gramme par mètre-cube ($\mu\text{g} / \text{m}^3$), puis de 20 à 40, etc., et on compte combien de fois on est tombé dans chaque classe. En divisant par le nombre total d'occurrences, on obtient un histogramme, que l'on appelle "loi de probabilité". Cette approche est très simple, très intuitive, et sera comprise par tous les élèves. Voici à quoi peut ressembler cette loi :



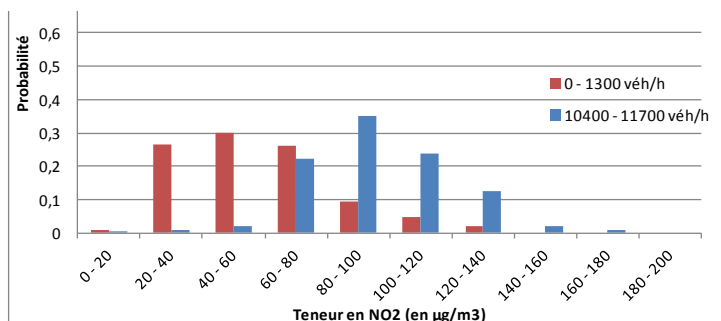
En quoi est-ce une loi "de probabilité", alors qu'elle provient de données parfaitement déterministes (aux erreurs de mesure près, bien sûr) ? En ce sens qu'elle caractérise la probabilité qu'aurait un visiteur (un Martien par exemple), s'approchant du Boulevard Périphérique à un instant quelconque (choisi de manière aléatoire avec équiprobabilité), de tomber sur telle ou telle valeur de pollution. Par exemple, la probabilité de tomber sur une pollution en NO₂ comprise entre 80 et 100 micro-gramme par mètre-cube est de l'ordre de 15 %.

Comme très peu de gens s'intéressent aux préoccupations des Martiens visitant le Périphérique, il aurait été plus parlant de dire que nous avons réalisé une présentation de l'information sous forme d'histogramme.

De manière générale, on aura une approche "probabiliste" si on ne peut pas, ou si on ne veut pas, chercher à connaître complètement les phénomènes qui sous-tendent l'expérience. Cela ne signifie en rien qu'elle soit régie par le hasard, mais que nous avons décidé de faire comme si, par choix ou par nécessité.

Ici, pollution ou pas, l'utilisateur respire un grand coup : "Ouf, je vois beaucoup mieux les tenants et les aboutissants, et je vois toutes les applications possibles. En effet, dans la réalité, on ne dispose jamais de toutes les informations nécessaires, et une approche probabiliste sera évidemment utile".

Notre expérience de base a commencé avec la loi de probabilité de la pollution en NO₂, indépendamment de tout le reste. Mais nous avons enregistré un autre paramètre, à savoir le trafic (mesuré en véhicules par heure). Nous allons donc pouvoir construire séparément les lois de probabilité de la pollution lorsque le trafic est faible et lorsqu'il est fort, et comparer ces deux lois. Voici un exemple de ce que l'on obtient, dans deux cas extrêmes (trafic très faible et très fort).



On pourrait évidemment réaliser ceci en choisissant une classe particulière de trafic, par exemple 2 000 à 2 500 véhicules par heure.

Nous venons de construire ici ce que l'on appelle des "lois de probabilité conditionnelles", terme impropre et difficile à comprendre. Il vaudrait mieux dire "lois de probabilité spécifiques", parce qu'elles sont spécifiques à une situation particulière. Concrètement, elles sont très faciles à calculer : on fait un tri informatique, ne gardant dans la situation générale que le cas particulier qui nous intéresse. Par exemple, on ne gardera que les situations où le trafic est $\leq 1\,300$ véhicules par heure, et on refait le calcul de la loi de probabilité de la pollution en NO₂ dans ce cas seulement. L'élève comprend cela immédiatement.

On peut ensuite s'interroger sur la notion d'indépendance : y a-t-il indépendance entre trafic et pollution : évidemment non ; comme on le voit sur le graphique ci-dessus, la loi de probabilité de la pollution se décale vers la droite lorsque le trafic augmente. Là encore, chacun comprend cela immédiatement.

On peut aussi multiplier les paramètres, chercher à caractériser leur importance respective, les classer, les hiérarchiser, etc.

Nous avons vu ici ce qu'on pourrait appeler "probabilité descriptive" : une expérience de la vie réelle est complexe ; le résultat n'est pas toujours le même ; il dépend de nombreux paramètres, plus ou moins bien connus, plus ou moins bien mesurés. La description probabiliste permet une approche de cette complexité.

Une autre situation que l'on rencontre souvent est celle de la "probabilité prédictive". Typiquement, elle concerne les pannes d'un appareil : 100 000 automobiles identiques ont été fabriquées ; combien seront encore en état de fonctionnement au bout de 5 ans ? La théorie sous-jacente est nettement plus difficile, et ne peut être présentée aux élèves que dans un deuxième temps, une fois que les probabilités descriptives ont été bien assimilées. Dans la question de la prévision, on introduit un "taux de risque", qui est la probabilité que le véhicule tombe en panne pendant une période donnée (mettons un an).

Ce taux de risque est inconnu, et il sera donc traité comme une variable aléatoire. On dispose d'un retour d'expérience : nombre de véhicules tombés en panne, selon l'âge, pendant chaque année, et il s'agit d'estimer ce taux de risque, en fonction des données disponibles. Même si la théorie est difficile, les hypothèses et les conclusions seront très claires pour les élèves.

Le premier exemple est facile ; le second l'est moins. De manière générale, la Théorie des Probabilités est difficile : elle décrit des lois de la Nature et ces lois sont toujours complexes. Je renvoie ici à l'exposé "Paradoxes Probabilistes" que j'ai fait en février dernier [2].

Nos recommandations

Il faut sortir d'un enseignement des probabilités purement axiomatique, et le faire reposer sur des exemples concrets ; l'apport général des méthodes probabilistes est celui de l'information partielle. On a un phénomène, inconnu ou mal connu, avec une série d'expériences, qu'en déduit-on ? Chaque enseignant pourra multiplier les exemples à sa guise.

Nous nous associerons volontiers (et bon nombre d'entreprises avec nous) aux efforts qui seront faits pour rendre plus accessible l'enseignement des probabilités. On peut facilement construire un site web, où chacun déposera ses exemples et viendra chercher ceux des autres, posera ses questions et apportera ses réponses, etc. On peut aussi penser à un bulletin de liaison (papier ou électronique), à l'organisation de colloques, de séminaires, etc. Nous sommes ouverts à toute suggestion en ce sens.

Références

[1] Liens entre trafic et pollution sur le boulevard périphérique autour de Paris ; résumé de l'étude réalisée pour la DGEC en 2015.

http://www.scmsa.eu/archives/SCM_resume_DGEC_2016_01_25.pdf

[2] Bernard Beauzamy : "Paradoxes probabilistes", exposé du 16 février 2016.

http://www.scmsa.eu/archives/BB_paradoxes_probabilistes_2016_02.pdf