Société de Calcul Mathématique SA

Outils d'aide à la décision

depuis 1995

∫

Phénomènes climatiques extrêmes

L'apport des méthodes probabilistes pour l'aide à la décision

par Cherif Seddik, Lucas Busson, Bernard Beauzamy

septembre 2024

Résumé opérationnel

La ville de Villiers le Bâcle, commune de l'Essonne, a été, il y a quelques années, victime d'orages spectaculaires : le réseau d'évacuation des eaux a été saturé et la commune partiellement inondée. Elle se demande : une telle situation est-elle exceptionnelle, ou bien devons-nous prévoir une amélioration du réseau ? Les méthodes probabilistes, essentiellement dues à Laplace, permettent de répondre à la question.

La difficulté principale, rarement traitée de manière satisfaisante dans la littérature, est la suivante : comment comptabiliser les tempêtes ? Nous ne pouvons pas nous limiter à compter celles qui ont frappé la ville elle-même : c'est réducteur. Si les tempêtes Lothar et Martin revenaient, elles pourraient passer 20 km au nord ou 30 km au sud. Nous considérons donc une quinzaine de villes dans le voisinage (rayon 30 km). Nous ne pouvons pas non plus affecter à la ville-cible toutes les tempêtes qui ont affecté le voisinage : ce serait se pénaliser de manière absurde, une tempête étant généralement localisée. Nous déterminons donc le risque pour chaque ville du voisinage, sur la base de son historique propre, en utilisant la méthode de Laplace. Enfin, nous affectons à la ville-cible le risque maximum, sur la base des 16 calculs qui ont été faits.

Tous calculs faits, notre conclusion a été : la probabilité qu'il ne se passe rien sur 5 ans est forte (0.7), donc pas d'urgence ; par contre, sur 10 ans, elle tombe à 0.5 ; on ne peut pas se permettre de ne rien faire.

I. Introduction

L'apport des méthodes probabilistes, en ce qui concerne l'aide à la décision, est à juste titre controversé : elles ne permettent pas de savoir si une tempête reviendra la semaine prochaine, l'an prochain, ou jamais. Les critiques les plus fermes ont été formulées par Claude Bernard, dans son livre "Introduction à la médecine expérimentale", 1865 :

"La statistique ne saurait donc enfanter que les sciences conjecturales ; elle ne produira jamais les sciences actives et expérimentales, c'est-à-dire les sciences qui règlent les phénomènes d'après les lois déterminées. On obtiendra par la statistique une conjecture avec une probabilité plus ou moins grande, sur un cas donné, mais jamais une certitude, jamais une détermination absolue."

A l'époque de Claude Bernard, les méthodes probabilistes n'en étaient qu'à leurs balbutiements. Nous pensons que lui-même, de nos jours, reconnaîtrait qu'elles apportent des renseignements indispensables :

- Pour un pays, dimensionner le nombre de lits nécessaires dans les cliniques et les hôpitaux, les stocks de médicaments, etc. ;

- Pour un individu donné, déterminer les maladies auxquelles il est le plus exposé, son espérance de vie, en fonction de ses caractéristiques (poids, âge, etc.) et de ses antécédents médicaux.

Chacun sait par exemple qu'un poids excessif prédispose aux maladies cardiovasculaires : cet énoncé est purement probabiliste et n'a rien de déterministe, mais il guidera le médecin dans le choix de ses examens et prescriptions.

II. La question posée

Elle ne porte pas sur une "prédisposition", mais appelle une réponse précise, immédiate et déterministe : oui ou non, devons-nous réviser notre réseau d'évacuation des eaux ? A cette question, nous avons apporté une réponse, également précise, immédiate et déterministe :

1) il n'y a pas d'urgence : la probabilité que quelque chose se produise en cinq ans est faible ;
2) vous devez néanmoins planifier l'amélioration du réseau : la probabilité sur dix ans est significative.

Beaucoup trouveront cette réponse incompréhensible : une période de dix ans est la réunion de deux de cinq, et si chaque période de cinq est tranquille, il doit en être de même pour celle de dix.

Eh bien non ! les lois probabilistes ont leurs paradoxes. Si vous jouez deux fois à pile ou face, la probabilité de sortir 2 P est 1/4, mais si vous jouez 4 fois, la probabilité de sortir au moins 2 P est 11/16. Elle était faible ; elle devient grande.

III. Méthodes probabilistes

L'énoncé fondamental est dû à Laplace ("Théorie analytique des probabilités", 1812) ; il se présente sous la forme d'une formule, caractérisant la probabilité qu'un événement ("accident") se produise n' fois sur N' essais futurs, sachant qu'il s'est produit n fois sur N essais passés. C'est ce qu'on appelle une "probabilité conditionnelle" : on dispose de l'information "n accidents sur N essais passés". Il est bien évident que l'information relative au passé a une importance : la probabilité future ne sera pas la même, s'il ne s'est rien passé auparavant ou s'il y a eu, au contraire, beaucoup d'accidents.

A. Enoncé

La formule de Laplace est :

$$p(n', N'; n, N) = \frac{N+1}{N+N'+1} \frac{\binom{N}{n}\binom{N'}{n'}}{\binom{N+N'}{n+n'}}$$

où $\binom{N}{n} = \frac{N!}{(N-n)!n!}$ avec la convention $0! = 1$. La démonstration de la formule n'est pas simple ; elle peut être trouvée dans le livre [MPPR]. Elle procède de l'idée suivante :

Si la probabilité de l'accident (on dit "du risque") était connue ; notons-la λ, la probabilité qu'il se produise n' fois en N' essais serait $p = \binom{N'}{n'} \lambda^{n'} (1-\lambda)^{N'} - n'$: c'est simplement la loi binomiale. L'information préliminaire "n accidents sur N essais" conduit à une certaine forme de loi de probabilité pour λ, d'où la formule découle. L'idée de Laplace, extrêmement ingénieuse, consiste à dire : nous ne connaissons pas la valeur exacte de λ (appelé "taux de risque"), considérons donc λ comme une variable aléatoire, dont nous allons préciser la loi en nous servant de l'historique.

Dans le cadre qui nous intéresse ici, les nombres N, N' seront exprimés en jours et n, n' le nombre de fois où l'on a vu une tempête pendant la période donnée. On fixe un seuil de pluviométrie, disons 80 mm d'eau en 24 h. L'énoncé $n = 2, N = 365$ signifie donc que l'on a vu deux jours où la pluviométrie a dépassé 80 mm en une année.

B. Hypothèses de travail

Comme toujours en probabilités, tout théorème requiert des hypothèses. Ici, il y en a deux :

- Les occurrences doivent être indépendantes. C'est le cas ici : si une tempête se produit, cela ne donne aucune indication sur la probabilité qu'elle revienne le jour suivant ou le mois suivant (ce serait différent pour un séisme, généralement suivi de répliques).

- La loi doit être stationnaire. C'est le cas ici, sur les périodes qui nous occupent : recul de 20 ans, anticipation de 10. Beaucoup de gens pensent que le climat est "détraqué", mais les statistiques sur les occurrences de phénomènes climatiques extrêmes montrent que la fréquence ne varie pas sur 50 ans, durée des observations. Pour des données détaillées, voir
  https://www.scmsa.eu/archives/SCM_Cyclones_2023_04_17.pdf

On dispose donc d'une formule, parfaitement bien établie, dont les hypothèses d'utilisation sont satisfaites. La question est dans la manière de l'utiliser.

IV. Difficultés d'utilisation

La formule est parfaitement simple et claire, s'il s'agit de compter le nombre d'objets imparfaits sur une chaîne de montage. Mais ici, il s'agit de tempêtes, et la question est : que compte-t-on ?

Si on se limite à la ville elle-même, on sous-estime le risque : si les tempêtes Lothar et Martin revenaient, rien ne dit qu'elles suivraient exactement la même trajectoire ; elles pourraient aussi bien passer 20 km au nord ou 30 km au sud.

Nous allons donc considérer un voisinage de la ville-cible (notée V_0), mettons un cercle de 30 km autour de cette ville. Nous dénombrons une quinzaine de villes équipées de pluviomètres dans ce voisinage ; notons-les $V_1,...,V_{15}$.

On ne peut pas attribuer à V_0 toutes les tempêtes ayant affecté le voisinage : ce serait surestimer grandement le risque. En effet, une tempête affecte une zone très réduite ; on constate sur les relevés que les tempêtes ayant affecté V_1 n'ont généralement pas affecté V_2, etc.

Ce sujet est généralement mal traité dans la littérature. Ici, nous adoptons l'approche suivante : nous cherchons, parmi les 16 villes $V_0,V_1,...,V_{15}$ celle qui est la plus exposée aux intempéries et nous affectons le risque à la ville V_0.

En d'autres termes :

- Nous tenons compte du voisinage : l'omettre serait absurde ;
- Nous ne faisons pas la somme des risques, le cumul serait erroné ;
- Nous affectons à la ville-cible le risque maximal parmi toutes les villes du voisinage.

Cela revient à dire que nous appliquons 16 fois la formule de Laplace, avec :

- Pour chaque ville, n, N donnés par l'historique (la profondeur de l'historique peut varier d'une ville à l'autre ; c'est généralement 20 ans = 7 300 jours).

- Pour chaque ville, $n' = 0$ (pas d'ennuis) et $N' = 1825$ jours (soit 5 ans) ou $N' = 3\,650$ jours, soit dix ans.

On obtient ainsi la probabilité de "pas d'ennuis" sur 5 ou 10 ans ; on prend la valeur la plus basse parmi les 16 villes et on affecte $1-p$ (probabilité d'avoir au moins un ennui pendant la période donnée) à la ville-cible.

Tous calculs faits, on trouve que la probabilité que quelque chose se produise (pluviométrie supérieure à 80 mm) est faible sur cinq ans : 0.3, mais significative sur dix ans : 0.48, d'où la conclusion : on peut prendre son temps, mais on ne peut pas se permettre de ne rien faire.