

6 - MODÉLISATION EN PROBABILITÉS CONDITIONNELLES

Michel HENRY

A - Situations de probabilités conditionnelles

A partir des issues **d'une expérience aléatoire**, on définit un univers Ω , ensemble des éventualités ω . On s'intéresse à deux événements A et B , parties de Ω , représentatives de deux caractéristiques ou propriétés possibles A et B de ces issues. Quand $\omega \in A$, on dit que « l'événement A est réalisé ». (Double statut du mot « événement ».)

On désire avoir une idée sur le lien entre A et B . Il peut être de nature logique : quand A est réalisé, alors B l'est aussi. Dans ce cas, on a simplement l'inclusion $A \subset B$ pour exprimer le fait que si $\omega \in A$, alors $\omega \in B$. On a : $P(A) \leq P(B)$.

Mais ce lien peut être de nature stochastique : quand on répète l'expérience, les événements A et B sont plus ou moins fréquemment réalisés conjointement, de telle sorte que lorsque l'on sait que l'un des deux est réalisé, cela donne une idée de nature probabiliste sur la réalisation éventuelle de l'autre. Quand cette fréquence est élevée, on parle de « forte corrélation », ou de dépendance stochastique. C'est cette notion qu'il s'agit de modéliser.

Dans la pratique, diverses situations relèvent des probabilités conditionnelles :

Les situations chronologiques

Elles sont particulièrement adaptées aux représentations avec des arbres, où l'expérience est décomposée en deux épreuves successives, les issues de la première conditionnent le déroulement (donc la répartition de la probabilité) de la deuxième.

Les situations causalistes

Elles s'apparentent aux précédentes et sont aussi assez bien décrites par des schémas en arbres. Voici un exemple type :

100 cobayes sont traités par trois produits provoquant une maladie M .

50 sont traités par le produit P_1 qui donne M avec la probabilité 0,25.

25 sont traités par le produit P_2 qui donne M avec la probabilité 0,5.

25 sont traités par le produit P_3 qui donne M avec la probabilité 0,3.

Un cobaye **pris au hasard** a la maladie M . Quelle est la probabilité qu'il ait été traité par P_1 ?

On y rencontre les mêmes obstacles didactiques quand il s'agit « d'inverser l'arbre » pour expliciter les probabilités des causes, conditionnellement à l'observation d'une conséquence possible. C'est dans ce cadre que la notion de probabilité conditionnelle a été historiquement progressivement dégagée.

Assez spectaculaire, la recherche des probabilités des causes, simple sur le plan mathématique, n'est pas exempte d'obstacles épistémologiques difficiles et de comportements erronés. Peut-on par exemple l'appliquer aux phénomènes historiques non reproductibles ? C'est pourquoi la formule de Bayes qui donne ces probabilités a été exclue du programme de terminale. Laplace la donne en sixième principe, dans des hypothèses maladroites, le conduisant à un raisonnement pour le moins insuffisant, comme nous le verrons au paragraphe suivant.

Les situations ensemblistes

Elles s'apparentent directement au modèle mathématique standard : l'issue d'une expérience aléatoire est porteuse de deux caractères qui peuvent être stochastiquement liés.

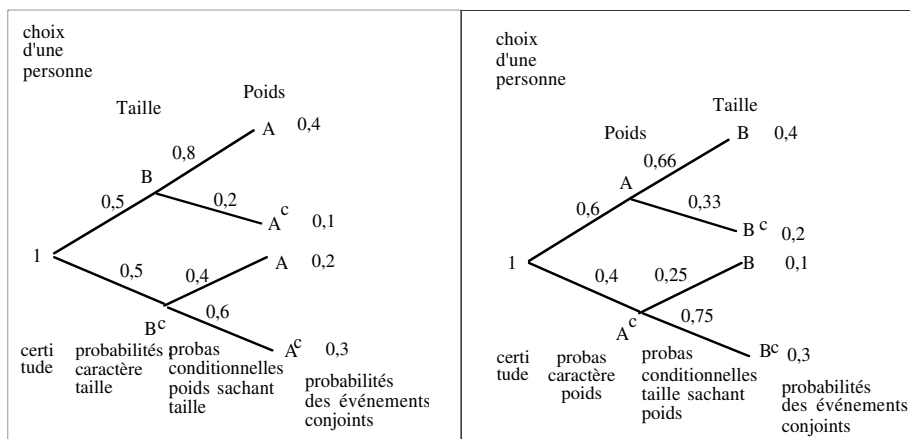
Par exemple dans un tirage au hasard d'une personne dans une population, l'observation croisée de son poids et de sa taille conduit à des résultats corrélés : les événements A = « la personne pèse plus de 65 kg » et B = « elle mesure plus de 1,75 m » sont sans doute liés.

Ces situations sont assez bien décrites par des tableaux de contingences à double entrée, donnant les probabilités conjointes et les

probabilités marginales des événements A et B étudiés. Elles peuvent aussi être décrites par des schémas d'arbres, dont l'enchaînement n'a pas d'ordre privilégié. Ils peuvent être plus facilement inversés que les précédents, d'un point de vue du sens.

Poids	Taille $B = \text{plus de } 1,75 \text{ m}$	$B^c = \text{moins de } 1,75 \text{ m}$	distribution marginale des probabilités des poids
$A = \text{plus de } 65 \text{ kg}$	0,4	0,2	0,6
$A^c = \text{moins de } 65 \text{ kg}$	0,1	0,3	0,4
distribution marginale des probabilités des tailles	0,5	0,5	1

Ce tableau donne les deux arbres suivants :



Cette notion de probabilité conditionnelle a été une source d'hésitations historiques. Voyons celles de D'Alembert et de Laplace.

B - Interprétations de D'Alembert et de Laplace

1 - D'Alembert et la navigation au XVIII^e siècle.

Dans l'article "*Probabilités*", de l'*Encyclopédie*, D'Alembert donne un exemple de *probabilité composée* :

*"Un ami est parti pour les Indes, sur une flotte de 12 vaisseaux, j'apprends qu'il en a péri trois, et que le tiers de l'équipage des vaisseaux sauvés est mort dans le voyage ; la probabilité que mon ami est sur un des vaisseaux arrivés à bon port, est 9/12, et celle qu'il n'est pas du tiers mort en route, est 2/3. La **probabilité composée** qu'il est encore en vie, sera donc les 2/3 de 9/12 ou 6/12 ou une demie certitude. Il est donc pour moi entre la vie et la mort".*

Dans cet exemple pseudo-concret, D'Alembert donne le principe de la multiplication des probabilités, comme on le ferait de proportions, sans dégager la notion de probabilité conditionnelle ni se préoccuper de l'indépendance des « événements », un peu tirés par les cheveux, qu'il examine.

2 - Les deux principes de Laplace

Pierre Simon LAPLACE est le mathématicien qui personnifie le tournant entre le XVIII^e et le XIX^e siècle, il peut être considéré avec Cauchy comme le maître à penser de l'école française qui étendra son influence jusqu'au XX^e siècle.

Il s'est intéressé très tôt aux probabilités, aussi bien du point de vue des applications que pour en faire une théorie mathématique à part entière, axiomatisée. Son premier ouvrage, publié en 1774, s'intitule : *Mémoire sur la probabilité des causes par les événements*. Il était nécessaire de faire la clarté sur les possibilités - et les limites- du calcul des probabilités dans ce domaine.

Dans sa présentation, il écrit :

"Je me propose de déterminer la probabilité des causes par les événements, matière neuve à bien des égards et qui mérite d'autant plus d'être cultivée que c'est principalement sous ce point de vue que la science des hasards peut être utile dans la vie civile".

Dans son *essai philosophique sur les probabilités* paru en 1825, Laplace rassemble les définitions et propriétés de base, énoncées en 10 "*principes*", qui peuvent être considérés comme une première axiomatique du calcul des probabilités.

Curieusement, Laplace énonce d'abord le sixième principe du calcul des probabilités des causes, formule abusivement appelée « de Bayes », avant de donner au septième principe la formule des probabilités totales :

Sixième principe :

“Chacune des causes, auxquelles un événement observé peut-être attribué, est indiquée avec d'autant plus de vraisemblance, qu'il est plus probable que cette cause étant supposée exister, l'événement aura lieu ;”

Traduction avec nos notations : soient C_1 et C_2 deux causes qui peuvent produire un événement E avec les probabilités respectives $P(E/C_1)$ et $P(E/C_2)$.

Si $P(E/C_1) > P(E/C_2)$, alors les probabilités des causes C_1 et C_2 sachant qu'on a observé E , sont dans le même ordre : $P(C_1/E) > P(C_2/E)$.

... “la probabilité de l'existence d'une quelconque de ces causes est donc une fraction dont le numérateur est la probabilité de l'événement, résultante de cette cause, et dont le dénominateur est la somme des probabilités semblables relatives à toutes les causes :”

Traduction avec nos notations : donc (!),

$$P(C_1 / E) = \frac{P(E / C_1)}{P(E / C_1) + P(E / C_2)} .$$

*... “si ces diverses causes considérées **a priori** sont inégalement probables, il faut au lieu de la probabilité de l'événement, résultante de chaque cause, employer le produit de cette probabilité, par la possibilité de la cause elle-même”.*

Traduction avec nos notations :

$$P(C_1 / E) = \frac{P(E / C_1)P(C_1)}{P(E / C_1)P(C_1) + P(E / C_2)P(C_2)} .$$

Laplace distingue ici deux notions : probabilité d'un événement et possibilité d'une cause, estimée a priori, sans que cette cause soit clairement issue d'une expérience aléatoire, comme le montre l'exemple de son urne qu'il traite ensuite. Dans cette interprétation causaliste, il lui est difficile de concevoir la probabilité de l'événement conjoint $P(E \cap C_1)$.

Septième principe :

“La probabilité d’un événement futur est la somme des produits de la probabilité de chaque cause, tirée de l’événement observé, par la probabilité que, cette cause existant, l’événement futur aura lieu.”

Traduisons : $P(E) = P(C_1)P(E/C_1) + P(C_2)P(E/C_2)$. C’est la formule des probabilités totales. Remarquons qu’avec la formule des probabilités composées : $P(A \cap B) = P(A/B).P(B)$, elle contient le sixième principe :

$$P(C_1 / E) = \frac{P(C_1 \cap E)}{P(E)} = \frac{P(E / C_1)P(C_1)}{P(E / C_1)P(C_1) + P(E / C_2)P(C_2)}.$$

3 - L’urne de Laplace

Laplace illustre alors ce principe des probabilités totales avec cet exercice classique de l’urne inconnue :

“Imaginons une urne qui ne renferme que deux boules dont chacune soit ou blanche, ou noire. On extrait une de ces boules, que l’on remet ensuite dans l’urne, pour procéder à un nouveau tirage. Supposons que dans les deux premiers tirages on ait amené des boules blanches ; on demande la probabilité d’amener encore une boule blanche au troisième tirage”.

Laplace tient alors le raisonnement suivant :

“On ne peut faire ici que ces deux hypothèses : ou l’une des boules est blanche, et l’autre noire, ou toutes deux sont blanches.

Dans la première hypothèse, la probabilité de l’événement observé est 1/4 ; elle est l’unité ou la certitude dans la seconde.

Ainsi, en regardant ces hypothèses comme autant de causes, on aura, par le sixième principe, 1/5 et 4/5 pour leurs probabilités respectives.

Or, si la première hypothèse a lieu, la probabilité d’extraire une boule blanche au troisième tirage est 1/2 : elle égale l’unité dans la seconde hypothèse ; en multipliant donc ces dernières probabilités, par celles des hypothèses correspondantes, la somme des probabilités, ou 9/10, sera la probabilité d’extraire une boule blanche au troisième tirage.”

Traduction : Les deux urnes possibles sont $C_1 = \{b ; n\}$ et $C_2 = \{b ; b\}$; deux tirages avec remises amènent l’événement $A = (b_1 ; b_2)$. Quelle est la probabilité de tirer encore une boule blanche au troisième coup (événement E) ?

Expression du septième principe :

$$P(E/A) = P(C_1/A)P(E/C_1 \text{ et } A) + P(C_2/A)P(E/C_2 \text{ et } A)$$

et on applique le sixième principe pour calculer $P(C_i/A)$:

$$P(C_1/A) = \frac{P(A/C_1)}{P(A/C_1) + P(A/C_2)} = \frac{\frac{1}{4}}{\frac{1}{4} + 1} = \frac{1}{5}$$

$$\text{et : } P(C_2/A) = \frac{P(A/C_2)}{P(A/C_1) + P(A/C_2)} = \frac{1}{\frac{1}{4} + 1} = \frac{4}{5}$$

$$\text{d'où : } P(E/A) = \frac{1}{2} \times \frac{1}{5} + 1 \times \frac{4}{5} = \frac{9}{10}$$

Dans ce calcul, Laplace fait implicitement l'hypothèse $P(C_1) = P(C_2)$.

Mais l'urne aurait pu être issue d'un choix aléatoire entre les trois possibilités suivantes : on lance un dé :

- si on obtient 1, on prend l'urne $\{n, n\}$ (C_0).
- si on obtient 2, 3 ou 4, on prend l'urne $\{b, n\}$ (C_1).
- si on obtient 5 ou 6, on prend l'urne $\{b, b\}$ (C_2).

Le calcul est alors différent :

$$P(C_1/A) = \frac{P(A/C_1)P(C_1)}{P(A/C_1)P(C_1) + P(A/C_2)P(C_2)} = \frac{\frac{1}{4} \times \frac{1}{2}}{\frac{1}{4} \times \frac{1}{2} + 1 \times \frac{1}{3}} = \frac{3}{11}$$

$$P(C_2/A) = \frac{P(A/C_2)P(C_2)}{P(A/C_1)P(C_1) + P(A/C_2)P(C_2)} = \frac{1 \times \frac{1}{3}}{\frac{1}{4} \times \frac{1}{2} + 1 \times \frac{1}{3}} = \frac{8}{11}$$

$$\text{d'où : } P(E/A) = \frac{1}{2} \times \frac{3}{11} + 1 \times \frac{8}{11} = \frac{19}{22} !$$

4 - Un exemple dans le style de Laplace.

Reprenons l'exercice des cobayes traités par les trois produits.
Comparons-le avec la situation suivante :

Dans la nature, il y a des rongeurs qui peuvent être porteurs de trois sortes de virus V_1 , V_2 , V_3 . Chaque virus donne la maladie M avec les probabilités 0,25 ; 0,5 ; 0,3.

On attrape un rongeur, il est malade. Quelle est la probabilité qu'il ait le virus V_1 ?

Le raisonnement de Laplace serait le suivant : n'ayant aucune indication sur la répartition des virus V_1 , V_2 , V_3 dans la région où on a attrapé le rongeur, je considère qu'il a été prélevé au hasard dans une population où l'on fait l'hypothèse d'équiprobabilité des trois sortes de virus :

$$P(V_1) = P(V_2) = P(V_3) = \frac{1}{3}.$$

C'est une hypothèse de modèle, sans doute fausse, mais qui semble la plus raisonnable.

On applique alors le sixième principe sous sa première forme :

$$P(V_1 / M) = \frac{P(M / V_1)}{P(M / V_1) + P(M / V_2) + P(M / V_3)} = \frac{0,25}{0,25 + 0,5 + 0,3} = 0,238$$

Mais quel sens a cette valeur numérique, avec de telles hypothèses de modèle, pour une « prise » de rongeur qui n'a rien d'une expérience aléatoire ?

C - Définition de la probabilité conditionnelle et modélisation

1 - Introduction du modèle

Comment interpréter le fait qu'une connaissance partielle sur un résultat d'une expérience aléatoire permet de modifier la probabilité que l'on va attribuer à un événement qui lui est lié ?

Il faut faire comprendre qu'une donnée nouvelle sur le déroulement d'une expérience aléatoire modifie non pas les probabilités des événements, mais la **description** de cette expérience, et par conséquent le modèle utilisé pour représenter les résultats possibles : on va changer de référentiel et déterminer une nouvelle répartition de probabilité sur le nouveau référentiel.

En fait, une approche fréquentiste va clarifier cette situation : **pour estimer une probabilité conditionnelle, on procède à une autre expérience aléatoire.**

Supposons en effet qu'à l'issue de l'expérience primitive, on sache que l'événement B est réalisé.

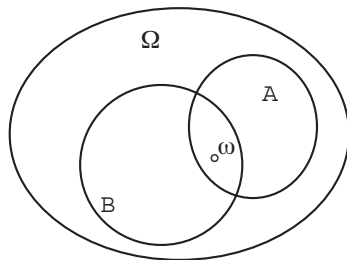
Reproduisons cette expérience un grand nombre de fois, en éliminant les épreuves où B n'est pas réalisé (on peut du moins imaginer cette répétition).

Cette procédure met en œuvre une autre expérience aléatoire : celle qui consiste à ne retenir l'issue de la précédente, **que** si B est réalisé, sinon on recommence.

Dans cette nouvelle expérimentation, la probabilité d'un événement A est déterminée par la fréquence des A intervenant dans les épreuves retenues. Elle est donc différente de la probabilité a priori de A , estimée par la fréquence des A intervenant dans l'ensemble des épreuves reproduisant l'expérience primitive.

2 - Définition ensembliste

La définition classique se plaçait résolument dans le cadre de la modélisation ensembliste, minimisant ainsi les obstacles dus à des conceptions subjectivistes. Reprenons ici cette définition, afin de montrer comment elle peut prendre en compte les remarques précédentes :



- Ω est l'ensemble référentiel représentant l'expérience aléatoire primitive.

- B est un événement de probabilité non nulle.

- A est l'événement dont on cherche la probabilité.

$\{\omega\}$ est l'événement élémentaire représentant l'issue de l'expérience.

Changement d'expérience : on sait que B est réalisé ($\omega \in B$). B est le nouveau référentiel. On désigne par P_B la probabilité des événements issus de cette expérience qui garantit la réalisation de B . A ne peut alors être réalisé que par la réalisation conjointe de A et B . La probabilité de l'événement A , $P_B(A)$, sera donc celle de $A \cap B$ avec comme donnée : $P_B(B) = 1$. On remarque

alors que si les événements élémentaires sont **équiprobables**, ils le sont aussi bien dans Ω que dans B et nécessairement, si $P(B) \neq 0$:

$$P_B(A) = \frac{\text{cas favorables}}{\text{cas possibles}} = \frac{\text{card}(A \cap B)}{\text{card } B} = \frac{P(A \cap B)}{P(B)}$$

3 - Interprétation fréquentiste

Mais si on est dans une situation de non-équiprobabilité, l'esprit fréquentiste du programme de terminale S enseigné en 2002 demande de revenir à l'introduction des probabilités par les fréquences.

Désignons par E l'expérience primitive représentée par le référentiel Ω , et B une partie de Ω représentant un événement B associé à E , de probabilité non nulle : $P(B) \neq 0$.

Si l'on reproduit E un grand nombre n de fois, la valeur observée f_B de la fréquence d'apparitions de B sera proche autant qu'on le veut de $P(B)$:

$$f_B = \frac{\text{nombre de réalisations de l'événement } B}{n}$$

Soit A un événement associé à E , on s'intéresse à la réalisation conjointe de A et B . $P(A \cap B)$ sera de même estimée par :

$$f_{A \cap B} = \frac{\text{nombre de réalisations conjointes de } A \text{ et } B}{n}$$

On s'intéresse à la fréquence d'apparition des A accompagnant la réalisation de B , celle-ci est donnée par :

$$f_{A/B} = \frac{\text{nombre de réalisations conjointes de } A \text{ et } B}{\text{nombre de réalisations de l'événement } B}$$

$$\text{d'où : } f_{A/B} = \frac{f_{A \cap B}}{f_B}$$

La probabilité d'obtenir A , lorsque l'on considère l'expérience aléatoire E_B consistant à n'accepter un résultat de E que lorsque B est réalisé, est estimée par : $f_{A/B}$.

La relation obtenue entre les fréquences conduit donc à la relation liant cette probabilité

$P_B(A)$ aux probabilités a priori de $A \cap B$ et de B :

$$P_B(A) = \frac{P(A \cap B)}{P(B)},$$

ce qui généralise cette même relation obtenue sous l'hypothèse d'équiprobabilité.

On remarque que $P_B(A)$ est proportionnelle à $P(A \cap B)$, ce qui est cohérent avec l'idée de réduire la probabilité de A à celle de $A \cap B$ lorsque B est réalisé, avec la contrainte : $P_B(B) = 1$.

A partir du sens ainsi donné à la notion de probabilité conditionnelle, les élèves peuvent aisément relier la définition de l'indépendance stochastique de A et de B , introduite dans le modèle sous la forme : $P(A \cap B) = P(A) P(B)$, avec l'idée concrète que l'on peut s'en faire :

A et B sont indépendants si la réalisation de B ne modifie pas la valeur de la probabilité de A d'un modèle à l'autre : $P_B(A) = P(A)$.

4 - Deux modèles mathématiques

L'expérience primitive E est représentée par l'univers Ω et l'espace probabilisé $(\Omega ; P_\Omega)$.

Soient A et B deux événements de E , représentés par les parties A et B de Ω de probabilités $P_\Omega(A)$ et $P_\Omega(B)$ non nulles.

On interprète le fait qu'une issue de E réalise B en introduisant une nouvelle expérience aléatoire E_B qui consiste à ne retenir E que si B est réalisé. E_B est représentée par l'univers B .

Les issues qui, réalisant B , réalisent A sont représentées par des ω , éléments de la partie $A \cap B$ de Ω . $A \cap B$ peut être considéré comme une partie de l'ensemble B , c'est la trace $t_B(A)$ de A sur B . La probabilité conditionnelle P_B est alors définie sur B par $P_B[t_B(A)]$, que l'on abrège en

$$P_B(A) = \frac{P_\Omega(A \cap B)}{P_\Omega(B)}.$$

Le nouvel espace probabilisé est alors $(B ; P_B)$.

Mais on peut aussi considérer que pour représenter la nouvelle expérience E_B , on conserve l'univers Ω ; la définition de la probabilité conditionnelle P_B revient simplement à changer la distribution de probabilité sur Ω , remplaçant P_Ω par P_B , définie sur Ω par :

$$P_B(A) = \frac{P_\Omega(A \cap B)}{P_\Omega(B)} .$$

Pour P_B , les parties du complémentaire de B sont de probabilités nulles.

Le nouvel espace probabilisé est dans ce cas $(\Omega ; P_B)$.

C'est généralement celui-ci que les probabilistes utilisent.

Notations

On peut rencontrer deux notations : $P_B(A)$ et $P(A/B)$. Le programme de terminale de 2002 impose $P_B(A)$.

La notation P_B signifie un changement de probabilité sur l'univers Ω , de la même manière qu'en analyse, on change de fonction.

La notation $P(A/B)$ fait plutôt penser à un changement d'univers : A/B serait la trace de A sur B et P apparaît plutôt comme l'abréviation de « probabilité » que comme le symbole d'une mesure sur B. Certains écrivent même : $\text{Prob}(A)$!

De plus l'ordre de lecture de A/B fait penser à un « événement » conditionnel. De nombreux exercices de terminale sont rédigés dans cette ambiguïté. Le concept d'événement conditionnel peut avoir un sens. Mais c'est une complication inutile, inductrice d'erreurs (on ne peut parler d'événements indépendants du type A/B et C/D , car B et D ne représentent pas la même expérience aléatoire).