Chapitre 21 : Probabilités

MPSI Lycée Camille Jullian

3 mai 2024

*Il est dans la probabilité que mille choses arrivent
qui sont contraires à la probabilité.*

Henri Louis Mencken

Aléatoire est dans l'univers.

Élève de lycée anonyme, en réponse à la première question

d'un exercice de probas posé par un collègue il y a quelques années.

Introduction via quelques exemples

Le concept de probabilité est a priori relativement intuitif : rien de suprenant à ce qu'un dé à six faces normalement constitué tombe en moyenne une fois sur six sur chacune de ses faces. Toutefois, ce tout petit exemple permet déjà de cerner quelques principes qui vont intervenir dans ce chapitre et qui rendent l'étude mathématique des probabilités un peu moins intuitive mais nettement plus riche qu'il n'y parait au premier abord. Déjà, il est essentiel de bien distinguer deux domaines des mathématiques qui sont proches mais qui exploitent des outils très différents : les probabilités et les statistiques. Les probabilités permettent simplement d'obtenir une estimation théorique de la fréquence d'apparition d'un évènement aléatoire, quand les statistiques étudient les fréquences réelles d'obtention de certains résultats lors d'expériences aléatoires (cf plus bas pour une définition claire des termes de vocabulaire). Ainsi, lorsque le probabiliste dit « Le dé va tomber sur le 6 une fois sur six », il ne s'agit bien entendu pas d'une prédiction qui va nécessairement se réaliser. Si on lance six fois de suite notre dé, il est tout à fait possible de ne jamais obtenir de 6, ou au contraire de l'obtenir plus d'une fois. L'étude de ce qui se passe « réellement » lorsqu'on lance notre dé relève de la statistique, que nous n'évoquerons pas vraiment cette année. Il existe toutefois des théorèmes importants qui relient les aspects statistique et probabiliste d'une expérience aléatoire. Par exemple, si vous répétez énormément de fois l'expérience consistant à lancer votre dé à six faces (une expérience virtuelle avec Python sera plus rapide à mettre en oeuvre et tout aussi efficace qu'un lancer de dé « à la main »), vous devriez constater que la fréquence d'apparition du 6 (nombre de fois où vous avez obtenu un 6 divisé par le nombre total d'essais) se rapproche de $\dfrac{1}{6}$. En fait, un théorème difficile connu sous le nom de loi faible des grands nombres assure que cette fréquence va effectivement **converger** vers la valeur théorique $\frac{1}{6}$, et donne même des estimations précises (faisant intervenir des probabilités !) sur l'écart entre la fréquence obtenue expérimentalement et la valeur théorique de la limite (du genre « pour n plus grand que 10 000, la probabilité que l'écart relatif entre la fréquence observée et $\frac{1}{6}$ soit supérieure à 5% est inférieure à 1% », les chiffres donnés ici ne sont pas exacts, vous étudierez ce théorème l'an prochain). Ce type de théorème a d'ailleurs des utilisations particulièrement importantes dans la vie pratique, puisqu'ils permettent par exemple de justifier le principe des sondages et de l'échantillonage utilisé pour les effectuer (on sait très bien, pour une population de votants de quelques dizaines de millions, quelle valeur de n prendre pour qu'interroger n personnes permette par exemple d'avoir une estimation des pourcentages réels de votants pour chaque candidat à une erreur relative de 5% près, et cette valeur de n est relativement faible par rapport aux quelques dizaines de millions).

Autre principe important : le fait qu'il n'existe pas une seule façon de calculer des probabilités pour une expérience aléatoire donnée. Ici, on a supposé que notre dé avait une probabilité identique de tomber sur chaque face (autrement dit qu'il est parfaitement équilibré), mais en pratique il est fort possible de concevoir un dé pour lequel cet équilibre ne sera pas vérifié. Si on truque le dé de façon à ce qu'il tombe deux fois plus souvent sur le 6 que sur le 1, l'expérience restera la même mais les calculs différents (c'est le concept de loi de probabilité défini plus loin dans ce cours). Par ailleurs, les outils mathématiques nécessaires à l'étude des probabilités dépendent fortement de l'expérience réalisée et notamment de l'univers mathématique qui lui est associé (ensemble des résultats possibles de l'expérience). C'est ce que nous allons constater à l'aide de quelques exemples bien choisis.

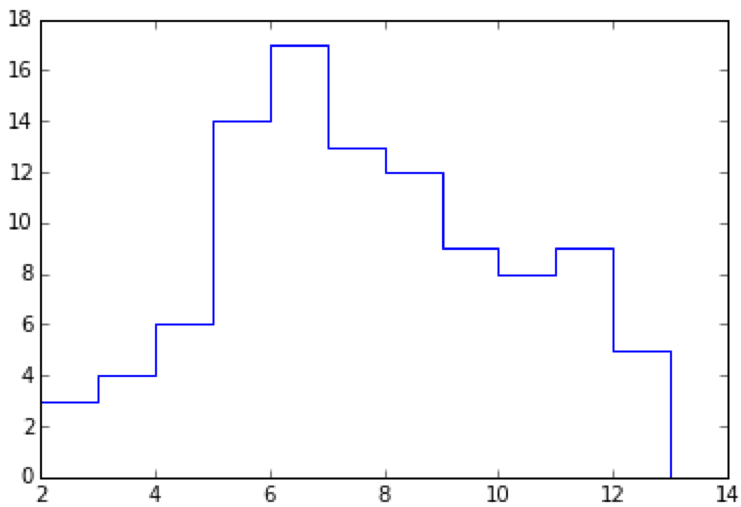
Premier exemple : probabilités finies.

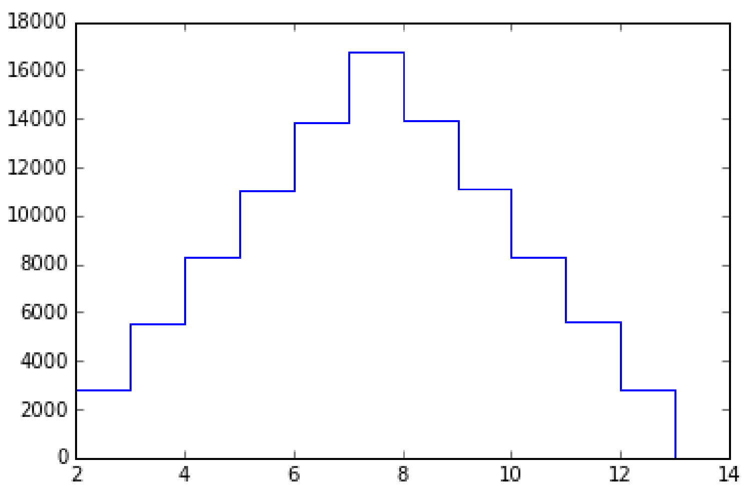
Prenons pour commencer un exemple simple où l'ensemble des résultats possibles est fini (d'où le terme de « probabilités finies ») : on lance simultanément deux dés à six faces (équilibrés) et on note la somme des deux résultats obtenus. Il est assez facile de se convaincre que tous les résultats possibles (en l'occurence les entiers compris entre 2 et 12) n'apparaitront pas avec la même fréquence, car il existe par exemple 4 façons d'obtenir une somme égale à 5 ($1+4$, $2+3$, $3+2$ et $4+1$), mais une seule d'obtenir 2 (les deux dés doivent tomber sur 1). Pour préciser cela, on peut considérer les choses de la façon suivante : il y a 6 résultats possibles pour le premier dé, autant pour le second, soit un total de 36 possibilités. On obtiendra donc une somme égale à 2 en moyenne une fois sur 36, mais une somme égale à 5 quatre fois sur 36, soit une fois sur 9. Cet exemple illustre bien la nécessité de bien définir quel est l'ensemble de résultats sur lequel on veut travailler, et surtout de vérifier si ces résultats sont équiprobables ou non. Pour s'amuser un peu avec cet exemple simple, on peut faire un peu de statistiques (ce sera la dernière fois de l'année !) en écrivant un petit programme Python qui effectue n simulations de cette expérience (pour un entier n choisi par l'utilisateur) et qui stocke dans une liste le nombre de fois que chaque résultat possible a été obtenu (on crée donc une liste contenant initialement 11 fois la valeur 0, puis après chaque expérience on augmente d'une unité l'élément de la liste correspondant à la somme obtenue, en décalant de façon à avoir le premier élément de la liste qui correspond au nombre de fois où on a obtenu 2, et le dernier qui compte le nombre de fois où on a obtenu 12). Tant qu'à faire, on utilise matplotlib pour nous tracer un petit histogramme des valeurs obtenues à la fin de l'expérience :

```
from random import randint
import matplotlib.pyplot as plt
def simulsomme(n) :
    l=[0 for i in range(11)]
    for in in range(n) :
        a=randint(1,6)+randint(1,6)
        l[a-2]+=1
    abscisse, ordonnee=[2],[0]
```

```
for i in range(11) :
        abscisse.append(i+2)
        ordonnee.append(l[i])
        abscisse.append(i+3)
        ordonnee.append(l[i])
plt.plot(abscisse+[13],ordonnee+[0])
return l
```

Si vous faites tourner le programme, vous constaterez sans surprise que, plus la valeur de n est grande (vous pouvez pousser Python assez loin, pour n de l'ordre d'un million, l'affichage ne prend que quelques secondes), plus l'histogramme ressemble à une pyramide bien équilibrée et symétrique, ce qui correspond bien entendu à la répartition théorique des différentes probabilités. Je donne pour le plaisir un exemple de ce qu'on obtient avec $n = 100$ (bien sûr, l'histogramme variera sensiblement si on répète exactement la même expérience) puis $n = 100\ 000$:





Sur cet exemple, on n'a pas vraiment eu besoin de calculs compliqués, mais on aurait déjà besoin de savoir calculer des sommes (et notamment les formules vues en début d'année pour certaines sommes classiques serviraient) si on voulait pousser jusqu'au calcul de l'espérance (moyenne) ou de la variance de la somme obtenue en lançant deux dés. Par exemple, le calcul de l'espérance revient à calculer la moyenne des résultats possibles, pondérés par leur probabilité d'apparition (il s'agit simplement d'un calcul de moyenne avec coefficients, sachant qu'ici les coefficients sont des probabilités dont par définition la somme est égale à 1, il n'y a donc pas besoin de diviser par quoi que ce soit). Ici, la moyenne m serait donc obtenue par le calcul $\displaystyle\sum_{i=2}^{12} i \times p_i$, en notant p_i la probabilité d'obtenir une somme égale à i (par exemple $p_5 = \dfrac{1}{9}$). On ne tombe pas vraiment sur une somme sympathique, mais comme il n'y a que onze termes, ça peut tout bêtement se calculer à la main (pour les curieux, on obtient logiquement une moyenne égale à 7, mais on verra des techniques plus subtiles pour retrouver ce résultat dans notre prochain chapitre de probabilités).

Surtout, les calculs vont très vite devenir nettement plus compliqués si on complexifie un tant soit peu le principe de l'expérience. Par exemple, si je vous demande la probabilité d'obtenir une somme égale à 12 en lançant simultanément cinq dés équilibrés à six faces, saurez-vous la calculer facilement ? Probablement pas, mais ce qui est certain, c'est que, si on souhaite vraiment se lancer dans ce genre de calcul, on aura besoin de bases solides dans un domaines des mathématiques que vous appréciez tout particulièrement : le dénombrement. D'ailleurs, ça tombe bien, puisque nous nous contenterons cette année d'étudier des probabilités finies. Allons tout de même jeter un oeil curieux à des cas plus complexes.

Deuxième exemple : probabilités discrètes.

On lance cette fois-ci un seul dé équilibré plusieurs fois de suite (on ne sait pas encore combien, vous allez vite comprendre pourquoi), jusqu'à ce que le dé tombe sur la face numéro 6. La situation est beaucoup plus compliquée puisqu'il y a ici une infinité de résultats possibles : un résultat où on ne tire qu'un fois le dé et on tombe sur 6, cinq résultats à deux lancers (1 puis 6, 2 puis 6 etc), 25 résultats à trois lancers etc... De plus, tous ces résultats ne sont certainement pas équiprobables puisque par exemple, obtenir 6 dès le premier lancer se produira avec probabilité $\dfrac{1}{6}$, alors qu'obtenir 2 puis 4 puis 6 aura une probabilité $\dfrac{1}{6^3} = \dfrac{1}{216}$ (il y a assez clairement des histoires de listes cachées derrière ces calculs de probabilités). Malgré tout, certains calculs restent faciles à effectuer : par exemple, déterminer la probabilité d'avoir besoin d'attendre exactement 10 lancers pour obtenir notre premier 6 revient à dire que chacun des neuf premiers lancers donne un autre résultat que 6, et que le dixième donne un 6, soit une probabilité de $\left(\dfrac{5}{6}\right)^9 \times \dfrac{1}{6}$. On aura en fait plus généralement une probabilité égale à $\dfrac{1}{6} \times \left(\dfrac{5}{6}\right)^{n-1}$ d'obtenir notre premier 6 au lancer numéro n, quel que soit l'entier naturel non nul n.

Mais on peut se poser des questions plus complexes à propos de cette expérience : est-il possible de ne jamais obtenir de 6, même après (expérience très théorique !) une infinité de lancers ? La réponse mathématique est un peu surprenante : oui, c'est possible, mais la probabilité que ça arrive est nulle ! En effet, la probabilité de ne toujours pas avoir obtenu de 6 après n lancers vaut $\left(\dfrac{5}{6}\right)^n$, et cette probabilité tend vers 0 quand n tend vers $+\infty$. Pourtant, il est bien sûr tout à fait possible d'écrire une liste infinie de nombres compris entre 1 et 5, ne faisant jamais apparaitre le 6 (il existe même une infinité de telles listes).

Autre question intéressante : combien de lancers faudra-t-il en moyenne pour obtenir notre premier 6 ? En adaptant le principe décrit plus haut, il faudrait ici calculer $m = \displaystyle\sum_{k=1}^{+\infty} k \left(\dfrac{5}{6}\right)^{k-1} \times \dfrac{1}{6}$. Mais oui, la somme est (nécessairement) devenue infinie puisqu'on a une infinité de valeurs possibles pour le nombre de lancers nécessaires avant de tomber sur un 6. Bon, ce genre de somme, qu'on étudiera un peu plus tard sous le nom de série, vous n'avez pour l'instant pas vraiment le droit de les écrire, et surtout vous ne disposez pas d'outils pour les manipuler simplement. Ici, écrire la somme infinie comme limite de sommes finies pourrait suffire à achever le calcul, mais le facteur k complique suffisamment les choses pour que je me contente de vous affirmer que cette moyenne est exactement égale à 6, ce qui revient à dire que si vous attendez quinze tours avant de pouvoir sortir votre premier cheval de l'écurie la prochaine fois que vous jouerez aux petits chevaux avec votre cousin de trois ans, c'est que vous n'avez vraiment pas de chance. Vous étudierez tout cela plus en détail l'an prochain.

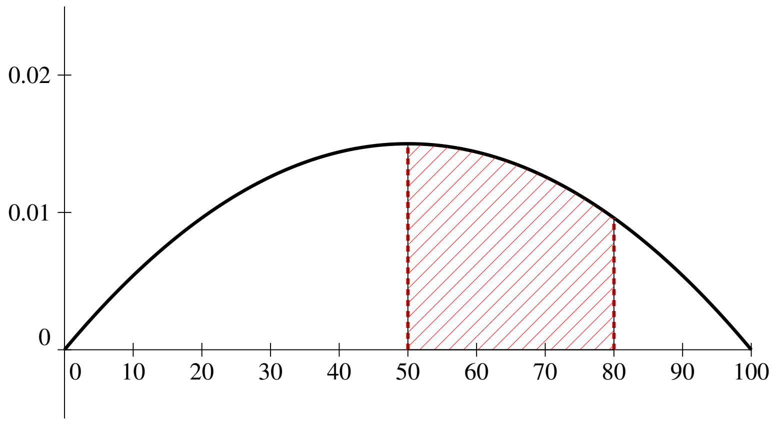
Troisième exemple : probabilités continues.

Une cible de jeu de fléchettes est constituée de trois disques concentriques de rayons respectifs 1, 2 et 3. Un joueur lance une fléchette dans la cible. On suppose (ce n'est pas très réaliste) que la fléchette atteint toujours la cible et que le point atteint dans la cible est aléatoire (avec une probabilité uniforme sur tous les points de la cible). On sort manifestement du cadre habituel : le nombre de résultats possibles n'est pas fini, loin s'en faut, mais il n'est même pas **discret** (c'est-à-dire qu'on ne peut pas « numéroter » toutes les possibilités à l'aides des entiers naturels, comme on l'avait fait dans l'exemple précédent). On peut tout de même attribuer de manière assez intuitive des probabilités à certains évènements : par exemple, il parait naturel de dire que la probabilité de tomber dans le disque central vaut un neuvième (rapport entre l'aire du disque central et celle de la cible). Mais que dire de l'événement « La fléchette tombe sur un point qui est à une distance rationnelle du centre »(oui, certes, personne ne se pose ce genre de question) ? Pas moyen de calculer facilement l'aire d'une telle chose. On admettra en fait que, dans ce genre de cas, on ne peut calculer les probabilités de tout et n'importe quoi, et qu'il faudra donc choisir quels sont les évènements autorisés. Ce cheminement est à l'origine du concept de tribu qui n'est pas à votre programme cette année (on ne s'intéressera qu'à de univers finis dans lesquels ce genre de problème ne peut pas se poser), mais sera peut-être évoqué dans votre cours l'an prochain. On note par ailleurs la première apparition dans un exemple d'expérience aléatoire de la notion d'aire, qui préfigure le lien extrêmement fort qui existe en fait entre les probabilités continues et une autre branche (a priori assez éloignée) des mathématiques : le calcul intégral. Et ça n'a en fait rien de surprenant si on songe que le calcul intégral n'est autre qu'un calcul de sommes « continues » : les probabilités finies donnent lieu à des calculs de sommes finies, les probabilités infinies discrètes à des calculs de sommes infinies discrètes (c'est exactement le principe d'une série), et les probabilités continues à des calculs de sommes continues, c'est-à-dire d'intégrales.

Quatrième exemple : probabilités continues.

On cherche à étudier une file d'attente (à la Poste, par exemple). À tout moment, il existe une certaine probabilité qu'une personne vienne s'ajouter à la file existante, et chaque personne passe au guichet un temps aléatoire. Ce temps est en pratique borné, mais peut prendre à peu près n'importe quelle valeur positive dans les limites du raisonnable (typiquement, admettons que si le temps d'attente est mesuré en minutes, il va appartenir à l'intervalle $[0, 100]$, mais n'a aucune raison d'être entier ou même rationnel, il peut prendre n'importe quelle valeur réelle entre 0 et 100). On pourrait naturellement décider de découper le temps en une multitude de petits intervalles (d'une seconde chacun, par exemple) et se ramener à des probabilités sur un ensemble fini, mais les calculs seraient affreusement lourds. Il est en fait plus logique d'accepter de faire des probabilités sur l'ensemble $[0, 100]$. Comme dans l'exemple précédent, on ne pourra bien sûr pas calculer la probabilité de n'importe quoi (par exemple la probabilité que le temps d'attente soit un nombre rationnel dont le dénominateur ne dépasse par 1 000), et surtout le calcul même des probabilités va faire intervenir de vrais calculs d'intégrale. On représente le temps d'attente par une fonction continue f définie sur l'intervalle $[0, 100]$, et dont l'intégrale entre 0 et 100 vaut 1 (cette intégrale représente en quelque sorte la probabilité totale). Pour déterminer la probabilité que le temps d'attente soit compris dans un intervalle $[a, b]$, on calcule alors simplement $\int_a^b f(t)\ dt$ (la fonction f est appelée fonction de densité de la variable aléatoire étudiée).

L'illustration ci-dessous donne un exemple possible de fonction de densité (je vous assure que l'intégrale entre 0 et 100 de la fonction est bien égale à 1), et l'aire de la zone hachurée représente dans ce cas la probabilité que notre temps d'attente soit compris entre 50 et 80 minutes (ici, la courbe de la fonction de densité est simple et symétrique par rapport à 50, mais ce n'est bien sûr pas du tout une obligation).



C'est par exemple le principe qui est derrière la définition des lois normales qui étaient encore au programme de maths du lycée avant la dernière réforme. Curieusement, vous ne ferez absolument pas de probabilités continues en prépa (ni cette année, ni l'an prochain), alors qu'il s'agit pourtant d'une très belle application du calcul intégral, et notamment des calculs d'intégrales impropres (intégrales à bornes infinies) que vous aborderez l'an prochain.

1 Vocabulaire

En probabilités, peut-être encore plus que dans tout autre domaine des mathématiques, la connaissance et le respect du vocabulaire technique sont essentiels, et ce d'autant plus que le vocabulaire en question est assez exotique. Vous devez donc parfaitement maitriser tous les termes introduits dans cette partie du cours.

1.1 Expérience aléatoire

Définition 1. Une **expérience aléatoire** est un phénomène ayant des résultats numériques dépendant du hasard.

Exemple : On lance un dé équilibré et on observe le résultat obtenu. On lance deux dés et on observe la somme des deux résultats obtenus. Bref, tous les exemples de l'introduction sont des expériences aléatoires.

Remarque 1. Insistons une fois de plus sur le fait qu'une expérience aléatoire est par définition non déterministe. L'étude des probabilités permet de faire des prévisions **statistiques** (qui seront en quelque sorte « de plus en plus valables » si on répète un grand nombre de fois l'expérience), mais en aucun cas de prévoir le résultat d'**une** expérience précise. Autrement dit, vous aurez beau être très forts en probas, ça ne vous aidera pas à décrocher la cagnotte au Loto. Par contre, ça peut vous aider à mieux jouer au poker (ou au bridge !).

Définition 2. On appelle **univers**, et on note Ω, l'ensemble des résultats possibles d'une expérience aléatoire.

Exemple : Dans notre premier exemple (on lance un seul dé), $\Omega = \{1, 2, 3, 4, 5, 6\}$. Dans le deuxième, c'est plus délicat : on peut considérer que $\Omega = \{2, \dots, 12\}$ puisqu'on s'intéresse à la somme des deux chiffres tirés, mais aussi que $\Omega = \{(1, 1), \dots, (6, 6)\}$ (tous les résultats possibles sur les deux dés). À vrai dire, peu importe le choix de l'univers tant qu'on effectue ensuite des calculs cohérents avec ce choix (en pratique, on choisira tout de même de préférence un univers dont tous les éléments sont équiprobables, donc ici on prendrait comme univers l'ensemble des couples d'entiers compris entre 1 et 6). Il arrivera d'ailleurs fréquemment que Ω soit beaucoup trop gros pour être décrit entièrement, connaître son cardinal sera souvent suffisant. Attention toutefois à ne pas confondre Ω, qui est un **ensemble**, et son cardinal, qui est un nombre entier. Le premier réflexe dans tout exercice de probabilités doit être de donner au moins le cardinal de l'ensemble Ω. C'est souvent pour la détermination de ce cardinal que le dénombrement interviendra (puisque ledit cardinal sera souvent un cardinal d'un ensemble de listes, d'arrangements ou de combinaisons d'un ensemble fini donné).

Remarque 2. Pour toute la suite de ce cours, on supposera que Ω est un ensemble fini.

1.2 Évènements

Définition 3. Un **évènement** (souvent noté A, B, ...) est un sous-ensemble de l'univers Ω. On dit qu'un évènement A est **réalisé** lorsqu'on effectue l'expérience si le résultat de l'expérience appartient à ce sous-ensemble.

Exemple : En pratique, un évènement est la plupart du temps décrit par une propriété plutôt que comme un sous-ensemble (ce qui serait de fait très lourd et peu lisible). Par abus de langage on dira ainsi qu'on considère par exemple l'évènement A : « on tire un double 5 »(dans le cas où l'expérience consiste à lancer deux dés) plutôt que de dire $A = \{(5, 5)\}$. Ce qui nous intéressera le plus en pratique sera le cardinal de l'ensemble correspondant (ici 1), qui permettra de calculer facilement la probabilité de l'évènement.

Définition 4. Il existe un vocabulaire précis pour certains évènements particuliers :

- L'évènement Ω tout entier est appelé **évènement certain**. C'est de fait un évènement qui se produira toujours.
- L'évènement vide est appelé **évènement impossible** et n'est jamais réalisé.
- Un évènement est **élémentaire** s'il est constitué d'un seul élément de Ω (c'était le cas de l'évènement A décrit dans l'exemple précédent, mais pas par exemple, pour la même expérience aléatoire, de l'évènement B : « Les deux dés ont donné le même résultat »).
- Deux évènements sont **incompatibles** si leur intersection est vide (autrement dit, ils ne peuvent pas être réalisés simultanément).
- Un **système complet d'évènements** est un ensemble d'évènements deux à deux incompatibles, et dont la réunion vaut Ω (autrement dit, une partition de Ω : on « découpe » l'ensemble de tous les résultats possibles en plusieurs morceaux qui ne peuvent pas se produire simultanément).

Exemples :

- L'évènement B : « La somme des deux chiffres donne 15 » est un évènement impossible.
- Les évènements C : « Le premier dé tombe sur 2 » et D : « On obtient une somme égale à 9 » sont incompatibles.

1.3 Lois de probabilité

Définition 5. Une **loi de probabilité** (ou plus simplement **probabilité** tout court) sur un univers Ω est une application $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$ vérifiant :

- $\mathbb{P}(\Omega) = 1$
- si $(A_1, A_2, \dots, A_n)$ est une liste d'événements deux à deux incompatibles inclus dans Ω, $\mathbb{P}\left(\bigcup_{i \in \{1,\dots,n\}} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ (propriété de σ-additivité).

Remarque 3. Une fois choisi l'univers correspondant à une variable aléatoire, le choix de la loi de probabilité correspond en fait à choisir quelles probabilités vont avoir les différents résultats de l'expérience. Ainsi, si l'expérience consiste à lancer un dé à six faces (pas forcément équilibré), le choix de la loi de probabilité revient à décider quelle sera la probabilité que le dé tombe sur 1, sur 2 etc. La première condition imposée revient simplement à dire que la somme de ces probabilités d'évènements élémentaires doit être égale à 1 (ce qui est intuitivement normal, on veut que la probabilité « totale » soit toujours égale à 1). La deuxième condition permet tout bêtement de reconstituer la probabilité d'un évènement à partir de probabilités d'évènements élémentaires. Ainsi, toujours en lançant un dé, si on souhaite connaitre la probabilité d'obtenir un chiffre pair, on additionnera les probabilités d'avoir obtenu un 2, un 4 ou un 6 (quels que soient les choix effectués pour les probabilités d'apparition de ces trois chiffres). Ces deux conditions suffisent en fait à garantir qu'on puisse calculer de façon cohérente la probabilité de n'importe quel évènement (et, on le verra, elles impliquent toutes les règles de calcul classiques sur les probabilités que vous connaissez déjà, comme la probabilité d'une union d'évènements pas nécessairement incompatibles, ou la probabilité d'un complémentaire). Assez curieusement, ces règles très simples sont même suffisantes quel que soit l'univers choisi (y compris si Ω est un ensemble infini).

Définition 6. Une **distribution de probabilités** sur l'ensemble fini Ω est une famille de réels $(p_\omega)_{\omega \in \Omega}$ appartenant tous à $[0, 1]$, indexée par l'ensemble Ω, et vérifiant $\sum_{\omega \in \Omega} p_\omega = 1$.

> **Proposition 1.** Si (p_ω) est une distribution de probabilités sur l'ensemble Ω, il existe une unique loi de probabilité $\mathbb{P} : \Omega \to [0, 1]$ telle que, $\forall \omega \in \Omega$, $\mathbb{P}(\{\omega\}) = p_\omega$.

Démonstration. Cela revient exactement à dire qu'une loi de probabilités est déterminée de façon unique par les probabilités des évènements élémentaires. À cause de la σ-additivité, on n'a de toute façon pas le choix : tout évènement A doit vérifier $\mathbb{P}(A) = \sum_{\omega \in A} p_\omega$. La vérification du fait que cette formule définit bien une probabilité sur Ω est complètement triviale. □

Définition 7. Un **espace probabilisé** est un couple $(\Omega, \mathbb{P})$, où $\mathbb{P}$ est une probabilité sur Ω.

Remarque 4. On peut très bien avoir envie de mettre plusieurs probabilités différentes sur un même univers. En reprenant l'exemple simple d'un lancer de dé, la probabilité « naturelle » consiste à décréter que $\mathbb{P}(1) = \mathbb{P}(2) = \cdots = \mathbb{P}(6) = \frac{1}{6}$, mais ce n'est bien sûr pas la seule ! Par exemple, $\mathbb{P}(1) = \frac{1}{21}$, $\mathbb{P}(2) = \frac{2}{21}$, …, $\mathbb{P}(6) = \frac{6}{21}$ définit également une probabilité, c'est même la seule à donner pour $\mathbb{P}(i)$ une valeur proportionnelle à i puisqu'il faut bien sûr vérifier que la somme des six probabilités élémentaires est égale à 1. En pratique, quand on résout un exercice de probabilités, la première chose à préciser, en plus de l'univers, est la loi de probabilité choisie sur cet univers (pour le transformer en espace probabilisé).

2 Propriétés

2.1 Généralités

Proposition 2. Si $\mathbb{P}$ est une loi de probabilité, on a toujours $\mathbb{P}(\emptyset) = 0$.

Démonstration. L'évènement vide étant incompatible avec lui-même (c'est bien le seul à vérifier cette curieuse propriété !), il doit vérifier $\mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset) = \mathbb{P}(\emptyset)$, donc $\mathbb{P}(\emptyset) = 0$. □

Remarque 5. Remarquez bien que la propriété $\mathbb{P}(\Omega) = 1$ fait partie de la définition d'une loi de probabilité, mais la propriété « symétrique » $\mathbb{P}(\emptyset) = 0$, elle, n'en fait pas partie, et découle de cette même définition.

Proposition 3. Pour tout évènement A, on a $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$.

Démonstration. Les évènements A et $\bar{A}$ sont incompatibles, donc on a $\mathbb{P}(A) + \mathbb{P}(\bar{A}) = \mathbb{P}(A \cup \bar{A}) = \mathbb{P}(\Omega) = 1$, ce qui donne bien le résultat voulu. □

Proposition 4. Si $A \subset B$, on a $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.

Démonstration. On peut écrire, de façon similaire à la démonstration précédente, $\mathbb{P}(A) + \mathbb{P}(B \backslash A) = \mathbb{P}(B)$. Or $\mathbb{P}(B \backslash A) \geqslant 0$ (par définition, une probabilité est toujours comprise entre 0 et 1), donc on a bien $\mathbb{P}(B) \geqslant \mathbb{P}(A)$. □

Proposition 5. Soient A et B deux évènements d'un même univers Ω, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Plus généralement, la formule de Poincaré est valable :

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} ((-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_k})).$$

Démonstration. On a par σ-additivité $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \backslash B)$; $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \backslash A)$ et $\mathbb{P}(A \cup B) = \mathbb{P}(A \backslash B) + \mathbb{P}(B \backslash A) + \mathbb{P}(A \cap B)$ (si vous n'êtes pas convaincus, rien n'empêche de faire un beau schéma à base de patates pour illustrer). L'égalité souhaitée en découle. Comme dans le cas des ensembles, on se gardera de faire une démonstration complète de la formule de Poincaré (qui se démontre d'ailleurs de la même façon que dans le cas des ensembles, par une récurrence assez laide). □

Proposition 6. Soit $(A_i)_{i\in\{1,\dots,n\}}$ un système complet d'évènements, alors $\sum_{i=1}^{n} \mathbb{P}(A_i) = 1$

De plus, quel que soit l'évènement B, $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)$.

Démonstration. Cela découle en fait immédiatement de la définition. Comme les A_i sont par définition incompatibles, $\mathbb{P}\left(\bigcup_{i\in\{1,\dots,n\}} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i)$. Or, la réunion des A_i vaut Ω (par définition également), donc sa probabilité vaut 1. La formule pour $\mathbb{P}(B)$ est similaire, il suffit de remarquer que les ensembles $B \cap A_i$ sont disjoints et que leur réunion est égale à B (en fait, ils forment un système complet d'événements pour B). Là encore, un dessin à base de patates est éclairant : si les évènements A_i forment un découpage de Ω en ensembles disjoints, alors B lui-même se découpe en sous-ensembles disjoints $B \cap A_i$. □

Remarque 6. Une formule similaire est valable dans le cas d'un système complet d'évènements infini, mais nous ne devrions pas en avoir besoin cette année.

2.2 Équiprobabilité

Définition 8. Il y a **équiprobabilité** sur l'espace Ω si tous les évènements élémentaires ont la même probabilité.

Proposition 7. Dans le cas de l'équiprobabilité, on a simplement,
$\forall A \in \mathcal{P}(\Omega)$, $\mathbb{P}(A) = \dfrac{|A|}{|\Omega|}$.

Démonstration. Posons $n = |\Omega|$. Les événements élémentaires formant un système complet d'événements, la somme de leur probabilités vaut 1. Or, cette somme est constituée de n nombres égaux (par définition de l'équiprobabilité), donc chacune de ces probabilités vaut $\dfrac{1}{n}$. Ensuite, un événement quelconque est union disjointe des évènements élémentaires qui le composent, sa probabilité vaut donc k fois $\dfrac{1}{n}$, où k est le nombre d'éléments dans cet événement. □

Remarque 7. C'est bien comme ça que vous calculez des probabilités depuis votre moyenne section de maternelle : nombre de cas favorables divisé par le nombre total de cas. Attention tout de même à bien comprendre que ce principe n'est justement valable que si tous les cas sont **équiprobables**.

3 Probabilités conditionnelles

Le principe des probabilités conditionnelles est, si on y réfléchit bien, assez simple, et surtout utilisé sans forcément qu'on s'en rende compte dans nombre d'exercices. Comme son nom l'indique, la notion désigne une probabilité soumise à une condition. Prenons un exemple simple : on lance deux dés (successivement, mais ça ne change rien aux probabilités) et on regarde leur somme (vous devez commencer à avoir l'habitude). On a déjà signalé que la probabilité d'obtenir 5 valait $\dfrac{1}{9}$. Supposons qu'on ait maintenant l'information supplémentaire suivante : on sait que le premier dé est tombé sur la face 2. Ça change tout ! Pour obtenir un total de 5, il faut maintenant (et il suffit) que le deuxième dé tombe sur 3, soit une chance sur 6. On dit que la probabilité d'obtenir 5 sachant que le premier dé a donné 2 vaut $\frac{1}{6}$ (naturellement, il sera plus commode de noter ceci à l'aide d'évènements). On a intuitivement une sentation de « cause et conséquence » quand on calcule une probabilité conditionnelle, qui traduit souvent la temporalité de l'expérience (ici, on connait le résultat du premier lancer de dé avant de lancer le deuxième), mais il faut bien avoir conscience que, mathématiquement, tout cela n'a guère d'importance. On peut calculer de la même façon la probabilité d'obtenir un certain résultat pour le premier dé en ayant une information sur le total des deux dés, ça ne changera rien !

3.1 Notations

Définition 9. Soit $(\Omega, \mathbb{P})$ un espace probabilisé et A, B deux évènements tels que $\mathbb{P}(A) \neq 0$. La **probabilité conditionnelle** de B sachant A est $\mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$.

Remarque 8. Si on veut être savant, on peut dire que l'application de $\mathcal{P}(\Omega)$ dans $[0,1]$ définie par $B \mapsto \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ définit une nouvelle probabilité sur Ω, appelée probabilité sachant A et notée $\mathbb{P}_A$.

Remarque 9. On rencontre souvent la notation alternative $\mathbb{P}(B\backslash A)$ pour la probabilité conditionnelle, mais nous ne l'utiliserons pas dans ce cours parce qu'elle entretient une confusion néfaste avec une probabilité de complémentaire.

Exemple : Cela correspond bien à l'idée intuitive. Sachant que A est réalisé, on se place dans un nouvel univers constitué des événements vérifiant A, et dans ce nouvel univers, B est réalisé pour tous les événements de $A \cap B$, ce qui conduit à la formule de la définition. Si on reprend l'exemple introductif, en notant A : « Le premier dé donne 2 » et B : « Le total vaut 5 », on a $\mathbb{P}(A) = \frac{1}{6}$ et $\mathbb{P}(A \cap B) = \frac{1}{36}$ (il n'y a qu'un cas qui marche, celui où on obtient 2 et 5), donc la probabilité conditionnelle vaut bien $\frac{1}{6}$. Si on préfère, on avait au départ $\Omega = \{(1,1), (1,2), \dots ; (6,6)\}$, ensemble de 36 possibilités parmi lesquelles quatre donnaient une somme égale à 5. Mais maintenant qu'on « sait » que le premier dé est tombé sur 2, l'univers des résultats possibles se réduit à $A = \{(2,1), (2,2), \dots, (2,6)\}$, nouvel univers au sein duquel une seule possibilité (la seule appartenant à $A \cap B$) donne une somme égale à 5.

Remarque 10. La probabilité conditionnelle étant une loi de probabilité, elle a les mêmes propriétés que n'importe quelle probabilité, en particulier $\mathbb{P}_A(\bar{B}) = 1 - \mathbb{P}_A(B)$, ou $\mathbb{P}_A(B \cup C) = \mathbb{P}_A(B) + \mathbb{P}_A(C) - \mathbb{P}_A(B \cap C)$. On peut donc effectuer des calculs compliqués de probabilités conditionnelles comme avec des probabilités « normales ».

Exemple : Un classique un peu perturbant. Une famille a deux enfants. On suppose que chaque enfant a une probabilité $\frac{1}{2}$ d'être un garçon, et donc $\frac{1}{2}$ aussi d'être une fille. On note A l'évènement « Les deux enfants de la famille sont des filles », B l'évènement « au moins l'un des deux enfants est une fille » et C l'évènement « l'aîné des deux enfants est une fille ». Les probabilités des trois évènements sont très faciles à calculer : $\mathbb{P}(A) = \frac{1}{4}$, $\mathbb{P}(B) = \frac{3}{4}$ (sur les quatre possibilités, il n'y a que la possibilité « garçon-garçon » où la famille n'a pas au moins une fille) et $\mathbb{P}(C) = \frac{1}{2}$. Mais si on calcule maintenant des probabilités conditionnelles, on constate que $\mathbb{P}_C(A) = \frac{1}{2}$ (c'est normal, si on sait que l'aîné est une fille, l'évènement A se réduit au fait que le deuxième enfant est aussi une fille). Par contre, $\mathbb{P}_B(A) = \frac{1}{3}$, la probabilité d'avoir deux filles sachant qu'on en a au moins une est donc inférieure à celle d'avoir deux filles sachant que l'aîné des enfants est une fille. Si ça vous parait logique, tant mieux. En tout cas c'est vrai !

3.2 Théorèmes

Les résultats énoncés dans ce paragraphe sont absolument fondamentaux, à comprendre et à maitriser parfaitement. Ils constituent tout simplement les outils de base pour faire des probabilités un peu plus évoluées que de simples calculs de dénombrement. En particulier, vous retrouverez extrêmement souvent la formule des probabilités totales dans les exercices cette année et l'an prochain.

Théorème 1. Formule des probabilités composées.

Soient $(A_1, A_2, \dots, A_n)$ des événements tels que $\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) \neq 0$, alors $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \times \mathbb{P}_{A_1}(A_2) \times \mathbb{P}_{A_1 \cap A_2}(A_3) \times \dots \times \mathbb{P}_{A_1 \cap A_2 \cap \dots \cap A_{n-1}}(A_n)$.

Remarque 11. La condition technique énoncée avant la conclusion du théorème sert simplement à assurer que toutes les probabilités conditionnelles sont bien définies : en effet, si $\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) \neq 0$, on a a fortiori $\mathbb{P}(A_1) \neq 0$; $\mathbb{P}(A_1 \cap A_2) \neq 0$, etc.

Remarque 12. Si on écrit la formule dans le cas où il n'y a que deux événements, on obtient simplement $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \times \mathbb{P}_{A_1}(A_2)$, ce qui est en fait la définition d'une probabilité conditionnelle.

Remarque 13. De façon plus générale, cette formule sert à formuler de façon rigoureuse le « principe de l'arbre » : quand une expérience aléatoire est consituée de plusieurs tirages (ou plusieurs étapes) successives, autrement dit quand on a une situation qu'on peut facilement représenter sous forme d'arbre de probabilités, la probabilité d'un évènement (visualisé comme un parcours suivant les branches de l'arbre) est égale au produit des différentes probabilités inscrites sur les branches de l'arbre « menant » à cet évènement (sur un arbre, ce sont bien des probabilités conditionnelles qui sont inscrites puisqu'on écrit sur chaque branche « secondaire » une probabilité qui suppose déjà vérifiée la condition correspondant à la branche « primaire » dont elle est issue).

Démonstration. On va procéder par récurrence. D'après la première remarque, toutes les probabilités conditionnelles sont bien définies, et d'après la seconde la formule est vérifiée pour $n = 2$. Supposons-la vraie au rang n, on a alors $\mathbb{P}\left(\bigcap_{i=1}^{n+1} A_i\right) = \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \times \mathbb{P}_{A_1 \cap A_2 \cap \dots \cap A_n}(A_{n+1})$ (on a simplement utilisé la formule dans le cas d'une intersection de deux événements), et via l'hypothèse de récurrence, on obtient immédiatement le résultat. □

Exemple : l'exemple bateau, genre de calcul qu'on fait en permanence sans même invoquer les probabilités composées (mais souvent en dessinant un arbre !). Dans une urne se trouvent 4 boules blanches et 3 noires. On tire successivement trois boules sans remise (et avec à chaque étape équiprobabilité de tous les tirages). La probabilité d'obtenir une noire, puis deux blanches vaut $\frac{3}{7} \times \frac{4}{6} \times \frac{3}{5} = \frac{6}{35}$.

Exemple : On dispose de trois enveloppes. La première contient 3 pièces d'un euro, la deuxième 5 pièces d'un euro et 3 de deux euros, la troisième 4 pièces d'un euro. On choisit une enveloppe au hasard, puis une pièce au hasard dans l'enveloppe (tout cela est bien entendu sous-entendu comme équiprobable à chaque étape). La probabilité d'obtenir une pièce de 2 euros vaut $\frac{1}{3} \times \frac{3}{8} = \frac{1}{8}$. Notons que le calcul de la probabilité d'obtenir une pièce d'un euro ferait intervenir la formule des probabilités totales (que nous allons énoncer maintenant) en plus de celle des probabilités composées (en gros car on n'a pas « une seule branche » de l'arbre menant à l'évènement dont on cherche à calculer la probabilité).

Théorème 2. Formule des probabilités totales.

Si les événements $(A_i)_{i\in\{1,\dots,n\}}$ forment un système complet d'événements et si $\forall i \in \{1,\dots,n\}$, $\mathbb{P}(A_i) \neq 0$, alors pour tout évènement B, on a $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i) \times \mathbb{P}_{A_i}(B)$.

Remarque 14. Dans le cas d'un système complet de deux évènements (ce qui revient simplement à dire que les deux évènements constituant le système complet sont complémentaires l'un de l'autre), on obtient la forme plus simple suivante, à retenir également sous cette forme : $\mathbb{P}(B) = \mathbb{P}(A) \times \mathbb{P}_A(B) + \mathbb{P}(\bar{A}) \times \mathbb{P}_{\bar{A}}(B)$.

Démonstration. Dans le cas particulier, on a vu plus haut que $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(\bar{A} \cap B)$. Or, on sait que $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}_A(B)$, et de même pour la deuxième moitié. Le cas général se fait exactement de la même manière. □

Remarque 15. Cette formule est à nouveau utile dans les cas de tirages à plusieurs étapes, donc dans les cas où on peut facilement représenter les choses sous forme d'arbre. Si on veut distinguer des cas selon le résultat obtenu lors de la première étape, on note $A_1, \dots, A_n$ les évènements correspondants à la réalisation des différentes possibilités à l'issue de cette première étape, puis on applique la formule des probabilités totales.

Exemple : Dans une urne se trouvent 4 boules noires et 6 blanches. On tire successivement trois boules dans l'urne, sachant qu'après chaque tirage on remet la boule tirée, mais qu'en plus on en ajoute une autre de la même couleur. La probabilité d'obtenir une boule noire au premier tirage vaut évidemment $\frac{2}{5}$. Si on veut calculer celle d'obtenir une boule noire au deuxième tirage, et qu'on veut rédiger les choses rigoureusement (sans se contenter de « lire un arbre »), on pose B_1 l'évènement « On a tiré une boule blanche au premier tirage » et N_1 l'évènement « On a tiré une boule noire au premier tirage ». On a bien sûr $\mathbb{P}(B_1) = \frac{3}{5}$ et $\mathbb{P}(N_1) = \frac{2}{5}$. Mais on peut aussi calculer facilement $\mathbb{P}_{B_1}(N_2) = \frac{4}{11}$ (si on a tiré une boule blanche lors du premier tirage, on a rajouté une boule blanche supplémentaire dans l'urne, et le deuxième tirage s'effectue donc avec sept boules blanches et quatre boules noires dans l'urne). De même, $\mathbb{P}_{N_1}(N_2) = \frac{5}{11}$ (cette fois, c'est une boule noire supplémentaire qu'on a ajoutée entre les deux tirages). Les évènements B_1 et N_1 formant un système complet d'évènements (c'est évident, ils sont complémentaires l'un de l'autre), on peut alors écrire la formule des probabilités totales appliquée à l'évènement N_2 : $\mathbb{P}(N_2) = \mathbb{P}(B_1) \times \mathbb{P}_{B_1}(N_2) + \mathbb{P}(N_1) \times \mathbb{P}_{N_1}(N_2) = \frac{3}{5} \times \frac{4}{11} + \frac{2}{5} \times \frac{5}{11} = \frac{22}{55} = \frac{2}{5}$. Autrement dit, c'est la même probabilité qu'au premier tirage. Au troisième tirage, elle vaudrait aussi $\frac{2}{5}$. Pas si évident que ça à justifier sans calcul ! Bien sûr, on peut trouver ces calculs évidents en faisant un arbre, mais il est important de s'entrainer à les rédiger correctement à l'aide de cette formule.

Exemple : Un type de problème très classique en probabilités et faisant intervenir les probabilités totales est celui des chaines de Markov (le terme n'est pas à retenir). Une chaine de Markov est une situation probabiliste qui évolue de façon discrète (c'est-à-dire qu'il se découle un intervalle de temps fixé entre chaque nouvelle évolution du système, par exemple une journée ou une minute), et dans laquelle chaque nouvel « état » dépend uniquement de l'état précédent du système et de probabilités conditionnelles. Par exemple, Homer Simpson mange tous les matins au petit déjeuner soit un donut, soit un cookie (voir plusieurs, le connaissant). On sait qu'au jour du début de l'expérience (numéroté jour 0), il a mangé un donut. On dispose par ailleurs des informations suivantes sur l'évolution du « système » constitué par les petits déjeuners d'Homer :

- s'il mange un donut au jour n, il mangera un cookie au jour $n+1$ avec probabilité $\frac{1}{2}$, et à nouveau un donut avec probabilité $\frac{1}{2}$.
- s'il mange un cookie au jour n, il passera au donut le lendemain avec probabilité $\frac{2}{3}$ et reprend un cookie avec probabilité $\frac{1}{3}$.
- il est donc sous-entendu qu'Homer mangera toujours soit un donut, soit un cookie, et jamais autre chose.

On cherche à déterminer en fonction de n la probabilité qu'Homer mange un donut au jour n. Notons donc d_n la probabilité qu'il mange un donut au jour n et c_n celle qu'il mange un cookie. On notera d'ailleurs de façon cohérente D_n l'évènement « Homer mange un donut au jour n » et C_n l'évènement « Homer mange un cookie au jour n ». On a par hypothèse $d_0 = 1$ et $c_0 = 0$, et les autres données de l'énoncé peuvent s'écrire sous forme de probabilités conditionnelles : $\mathbb{P}_{D_n}(D_{n+1}) = \mathbb{P}_{D_n}(C_{n+1}) = \frac{1}{2}$; $\mathbb{P}_{C_n}(C_{n+1}) = \frac{1}{3}$ et $\mathbb{P}_{C_n}(D_{n+1}) = \frac{2}{3}$. De plus (c'est essentiel pour que cela fonctionne), les évènements D_n et C_n (quelle que soit la valeur fixée de n) forment un système complet d'évènements (ils sont bien complémentaires vu les hypothèses faites sur les petits déjeuners d'Homer). On peut alors appliquer la formule des probabilités totales pour calculer $d_{n+1} = \mathbb{P}(D_{n+1}) = \mathbb{P}(D_n) \times \mathbb{P}_{D_n}(D_{n+1}) + \mathbb{P}(C_n) \times \mathbb{P}_{C_n}(D_{n+1}) = \frac{1}{2}d_n + \frac{2}{3}c_n$. On peut d'ailleurs obtenir de même une relation exprimant c_{n+1} en fonction de d_n et de c_n, mais elle ne nous servirait à rien pour achever les calculs.

Utilisons maintenant simplement le fait que $c_n = 1 - d_n$ (puisque D_n et C_n sont complémentaires) pour écrire $d_{n+1} = \frac{1}{2}d_n + \frac{2}{3}(1 - d_n) = \frac{2}{3} - \frac{1}{6}d_n$. La suite (d_n) est donc une suite arithmético-géométrique, d'équation de point fixe $x = \frac{2}{3} - \frac{1}{6}x$, ce qui donne $x = \frac{2}{3} \times \frac{6}{7} = \frac{4}{7}$. Posons donc $u_n = d_n - \frac{4}{7}$, on a alors $u_{n+1} = d_{n+1} - \frac{4}{7} = -\frac{1}{6}d_n + \frac{2}{3} - \frac{4}{7} = -\frac{1}{6}d_n + \frac{2}{21} = -\frac{1}{6}\left(d_n - \frac{4}{7}\right) = -\frac{1}{6}u_n$. La suite (u_n) est donc géométrique de raison $-\frac{1}{6}$ et de premier terme $u_0 = d_0 - \frac{4}{7} = \frac{3}{7}$. On en déduit que $d_n = u_n + \frac{4}{7} = \frac{4}{7} + \frac{3}{7}\left(-\frac{1}{6}\right)^n$. On pourrait bien sûr calculer de même c_n, ou encore plus rapidement obtenir sa valeur en exploitant la relation $c_n = 1 - d_n$.

Notons au passage que, quand n tend vers $+\infty$, la limite de la suite d_n vaut $\frac{4}{7}$. Autrement dit, à long terme, Homer mange des donuts en moyenne quatre jours par semaine. Dans ce genre de problème, les suites étudiées ont presque systématiquement une limite finie. Dans le cas d'une chaine de Markov à plus de deux états (par exemple on autorise Homer à manger un troisième type de nourriture pour son petit déjeuner), de simples calculs élémentaires sur les suites ne suffisent plus, et il faut utiliser du calcul matriciel pour s'en sortir. Nous verrons un tel exemple un peu plus tard dans le cours, mais la résolution du problème se ramènerait à peu de choses près au calcul des puissances d'une matrice carrée de taille 3. Et pas n'importe quelle matrice, mais une matrice stochastique (je vous laisse revoir la feuille d'exercices numéro 8 si vous voulez essayez de comprendre le lien).

Théorème 3. Formule de Bayes.

Soient A et B deux évènements de probabilités non nulles, alors $\mathbb{P}_B(A) = \dfrac{\mathbb{P}(A) \times \mathbb{P}_A(B)}{\mathbb{P}(B)}$.

Démonstration. Il n'y a presque rien à faire, on sait que $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}_A(B) = \mathbb{P}(B) \times \mathbb{P}_B(A)$, la formule en découle immédiatement. □

Remarque 16. On peut également donner une version de cette formule avec plus de deux événements, mais cela ne présente guère d'intérêt pratique. Cette formule, qui n'apporte en apparence pas grand chose par rapport aux précédentes, est en fait très utile dans la mesure où elle permet de « remonter le temps » lorsqu'on a une expérience faisant apparaitre des choix chronologiques. Autrement dit, lorsqu'on dispose de probabilités conditionnelles « naturelles », on peut calculer des probabilités conditionnelles « dans le mauvais sens ». Par exemple dans le cas d'une expérience aléatoire en deux étapes successives, on dispose d'une inforamtion sur ce qui s'est passé à la fin de l'expérience, et on souhaite en déduire la probabilité que quelque chose se soit produit lors de la première partie de l'expérience. Cf l'exemple qui suit.

Exemple : On peut reprendre n'importe quel exemple classique, mais en essayant de faire les choses dans le sens inhabituel. Pour faire simple, on tire deux dés successivement, la somme des deux chiffres obtenus vaut 9. Quelle est la probabilité que le premier dé soit tombé sur 4 ? Notons A l'évènement « Le premier dé tiré tombe sur 5 » et B l'évènement « La somme des deux chiffres obtenus vaut 9 ». On a $\mathbb{P}(A) = \dfrac{1}{6}$, $\mathbb{P}(B) = \dfrac{1}{9}$ et $\mathbb{P}_A(B) = \dfrac{1}{6}$, donc $\mathbb{P}_B(A) = \dfrac{\frac{1}{6} \times \frac{1}{6}}{\frac{1}{9}} = \dfrac{1}{4}$.

3.3 Indépendance d'événements

Encore une notion relativement intuitive a priori, mais qui nécessite une définition mathématique précise. On a envie de dire que deux événements sont indépendants si la réalisation de l'un n'influence pas celle de l'autre. Par exemple, en reprenant comme souvent notre lancer de deux dés, l'événement « Le premier dé tombe sur un chiffre pair » devrait logiquement être indépendant de l'événement « Le deuxième dé tombe sur 1 ou 2 ». Attention à ne surtout pas confondre cette notion avec la notion d'évènements incompatibles, qui n'a absolument rien à voir !

Définition 10. Deux événements A et B sont **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Proposition 8. Deux événements de probabilité non nulle sont indépendants si et seulement si $\mathbb{P}(B) = \mathbb{P}_A(B)$.

Démonstration. Rappelons que sous ces hypothèses $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}_A(B)$. En identifiant les deux formules, on obtient tout de suite le résultat. La deuxième formulation correspond exactement à la notion intuitive « la réalisation de A n'influence pas la réalisation de B ». □

Remarque 17. Dans le cas où l'un des deux événements a une probabilité nulle, les événements sont de toute façon nécessairement indépendants (même si c'est absurde !).

Exemple On tire une carte dans un jeu de 32 cartes. Les événements « obtenir un Roi » et « obtenir un pique » sont indépendants.

Remarquez que, dans le but de prouver l'indépendance de deux événements, les deux formulations sont équivalentes (aucune des deux n'est vraiment plus simple à manipuler que l'autre).

Définition 11. Des événements $(A_1, A_2, \dots A_n)$ sont dits **mutuellement indépendants** si $\forall I \subset \{1, 2, \dots, n\}$, on a $\mathbb{P}(\bigcap\limits_{i \in I} A_i) = \prod\limits_{i \in I} \mathbb{P}(A_i)$.

Remarque 18. Attention, des événements mutuellement indépendants sont indépendants deux à deux mais la réciproque n'est pas vraie ! La condition est beaucoup plus forte que ça. Par exemple, pour trois événements, on doit avoir $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) \times \mathbb{P}(A_2)$, $\mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_1) \times \mathbb{P}(A_3)$, $\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_2) \times \mathbb{P}(A_3)$ (indépendance deux à deux des trois évènements) mais aussi $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \times \mathbb{P}(A_2) \times \mathbb{P}(A_3)$ (ce qui ne découle pas des autres conditions). Les conditions à vérifier deviennent rapidement affreuses quand on augmente le nombre d'évènements.

Proposition 9. Si A et B sont indépendants, A et $\bar{B}$ le sont aussi.

Démonstration. C'est un petit calcul : $\mathbb{P}(\bar{A} \cap B) = \mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A) \times \mathbb{P}(B) = \mathbb{P}(B) \times (1 - \mathbb{P}(A)) = \mathbb{P}(B) \times \mathbb{P}(\bar{A})$. □

Remarque 19. Ce résultat se généralise à plus de deux évènements : si $(A_1, A_2, \dots, A_n)$ sont mutuellement indépendants, on peut remplacer une partie des A_i par leur complémentaire et conserver l'indépendance mutuelle.

Exemple : On fait une série de n lancers de pièces. Les événements A_k : « On obtient Pile au $k$ème lancer » sont mutuellement indépendants.

Exemple : On lance deux dés (si, si, je vous jure, encore une fois), et on considère les événements A : « Le premier dé donne un résultat pair », B : « Le deuxième dé donne un résultat pair » et C : « Les deux dés donnent des résultats de même parité ». Je vous laisse vérifier que $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ et $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}$, donc les événements sont deux à deux indépendants. Pourtant, $\mathbb{P}(A \cap B \cap C) = \frac{1}{4}$, donc les trois événements ne sont pas mutuellement indépendants.