Probabilités, Statistiques, Combinatoire

Philippe Duchon

16 janvier 2018

Ces notes sont le support du cours du même nom proposé en deuxième année de Licence Informatique (parcours Informatique et Mathématiques-Informatique) de l'université de Bordeaux.

Table des matières

Chapitre 1

Introduction

1.1 Pourquoi ce cours

Qu'est-ce que les probabilités ? Les statistiques ? Et que viennent-elles faire dans un cursus universitaire d'informatique ?

De telles questions sont pertinentes, et je vais essayer d'y répondre honnêtement.

Probabilités et statistiques sont deux domaines des mathématiques qui cherchent à traiter la situation extrêmement courante dans laquelle on doit faire face à l'incertitude. Cette incertitude peut avoir diverses causes ; les plus fréquentes sont sans doute le fait qu'on ne dispose que d'informations partielles, ou peu sûres (qu'on pense au responsable d'un institut de sondages en période électorale), ou encore qu'on cherche à anticiper le futur (savoir si le futur est totalement déterminé par le présent est autant une question philosophique que scientifique, mais à l'échelle où la plupart d'entre nous se situent, le résultat est le même : en pratique, de nombreux aspects du futur même proche sont inaccessibles à la prédiction).

Et pourtant, les situations, dans la vie de tous les jours, où l'on est amené à prendre des décisions malgré une bonne dose d'incertitude sont légion.

1.2 La question du hasard : théorie des probabilités

Qu'est-ce que le hasard ? Est-ce qu'il existe, seulement ? La question n'est pas tranchée, et la théorie des probabilités ne prétend pas le faire. Au niveau microscopique, quand on "lance un dé" (*i.e.* qu'on prend un cube approximativement homogène, dont les faces sont numérotées de 1 à 6, et qu'on le lâche dans une position et avec une vitesse plus ou moins arbitraires, en espérant que, sous l'effet des lois de la mécanique - notamment de la gravité, des frottements et des chocs - il s'immobilise assez rapidement en reposant sur une surface à peu près plane, avec une de ses faces en contact de la surface, ce qui permet de désigner la face opposée comme donnant "le résultat"), rien ne permet *a priori* de déclarer que le résultat soit choisi "au hasard", encore moins que chacun des entiers de 1 à 6 ait les mêmes chances d'apparaître.

Ce que prétend faire la partie des mathématiques que l'on nomme théorie des probabilités, c'est décrire un certain nombre de règles, d'*axiomes*, qui sont censés être valide pour toute expérience "aléatoire", et qui permettent, à partir d'un *modèle* que l'on se donne, de **faire des prédictions quantitatives**. En gros, on va, à partir de ces axiomes et de ce modèle, être capable de calculer un degré de vraisemblance (*probabilité*) à des événements incertains

qui peuvent avoir une description passablement complexe.

Cela permet, par exemple, de faire des prédictions de ce genre :

— si le professeur confisque, au début du cours, les téléphones de n étudiants et, à la fin du cours, les redistribue "au hasard" en rendant exactement un téléphone à chacun (et ici il y a un modèle, implicite : les téléphones sont "bien mélangés", *i.e.* chacune des $n!$ répartitions possibles des n téléphones entre les n étudiants a la même chance d'être choisie), est-il plausible que personne ne récupère son propre téléphone ? Un calcul complexe montre que, pour peu que n soit plus grand que quelques unités, cela devrait se produire avec une fréquence proche de $1/e \simeq 0.366$, soit un peu moins de 37%.

— si on lance de manière répétée un dé "équilibré" (modèle : chaque lancer a la même chance de donner chaque résultat possible, et les lancers sont "indépendants"), en se promettant de ne s'arrêter que lorsque chacun des 6 entiers de 1 à 6 aura été obtenu au moins une fois, est-on sûr de s'arrêter un jour ? (oui, du point de vue probabiliste) est-il plausible de s'arrêter au bout de 6 lancers ? (possible, mais peu probable ; 1.54%) En moyenne, combien de lancers doit-on s'attendre à faire ? (14.7)

— si l'on insère les entiers de 1 à 10, dans un ordre "au hasard" (modèle : les $10! = 3628800$ ordres d'insertion ont tous la même chance d'être choisis), dans un arbre binaire de recherche, quelle est la probabilité d'obtenir un arbre catastrophique de hauteur 9 ? (environ 0.00056, soit moins de 0.06%)

1.3 Interprétation des données : statistiques

Contrairement aux probabilités, qui partent d'un modèle et cherchent à faire des prédictions, les statistiques suivent un chemin radicalement différent : partant d'observations, elles cherchent à proposer un modèle le plus proche possible de la "réalité".

On se place donc dans la situation où un certain nombre de données ont été récoltées, par sondage ou par échantillonnage, et où l'on cherche à déterminer un modèle raisonnable pour ces données.

Par exemple, pour se faire une idée de la répartition, dans la population des personnes âgées de 20 à 25 ans (soit quelques millions de personnes, si l'on s'en tient à la population française), des tailles, ou des poids, ou des couples (taille, poids), ou, peut-être de manière plus pertinente, des triplets (taille, poids, sexe), on peut essayer de relever ces valeurs (ou paires de valeurs, ou triplets de valeurs) pour un échantillon "aléatoire" d'individus au sein de la population concerneée, et d'en inférer des valeurs plausibles, du type "il est plausible que le poids moyen des hommes d'entre 20 et 25 ans soit comprise entre X et Y kg", "si la répartition des tailles des femmes âgées de 20 à 25 ans suit une *courbe en cloche* avec une moyenne de m et un écart-type de s, alors les valeurs les plus plausibles pour m et s sont M et S".

De telles estimations reposent inévitablement sur des hypothèses (que la procédure d'échantillonnage soit suffisamment uniforme ; que, dans le cas de recueil d'opinions pour un sondage, les réponses fournies reflètent bien les opinions réelles des personnes ; que la fameuse "courbe en cloche" soit bien adaptée à la description des données en question), et les estimations sont généralement accompagnées [1] de "marges d'erreurs" indiquant le degré de précision qu'on

1. Ou du moins, elles *devraient* en être accompagnées ; l'observateur avisé des sondages publiés par la presse en période électorale aura sans doute remarqué que le chiffre estimé est toujours beaucoup mieux mis en avant

peut espérer leur accorder.

1.4 Pour l'informaticien

L'informaticien est un citoyen comme un autre : il a autant que les autres intérêt à comprendre comment déchiffrer un sondage, ou à savoir estimer un risque. Ces considérations, à elles seules, ne justifient pas complètement l'inclusion d'un cours de probabilités et statistiques dans un cursus d'informatique. En revanche, ce qui la justifie, c'est la multitude de situations où ces connaissances se révèlent utile, voire indispensables, à l'activité de l'informaticien, qu'il soit développeur, administrateur d'un système, chercheur, ...

— Lorsque l'on étudie la complexité d'un algorithme, il est parfois souhaitable d'aller au-delà de la classique "complexité dans le pire des cas". Un algorithme comme **QuickSort** a longtemps été l'algorithme de choix pour trier des données par comparaisons successives, alors même qu'il est facile de voir que sa complexité dans le pire des cas est mauvaise, $\Theta(n^2)$; ce qui le sauve et le rend attractif, c'est que l'on peut montrer que, si les données initiales sont "raisonnablement bien mélangées", il est *extrêmement probable* que ses performances soient "bonnes".

— Lorsque l'on cherche à dimensionner les ressources nécessaires à un système qui va fonctionner de manière décentralisée (penser à la mémoire qu'on installe sur un équipement réseau), on est naturellement amené à poser un modèle probabiliste pour la façon dont il va être sollicité (en moyenne tant de paquets par seconde, distribués dans le temps de telle ou telle manière), et à se poser des questions de nature probabiliste ("en fonctionnement normal, à partir de quelle quantité de mémoire puis-je considérer qu'il y a moins de 1% de chances que mon équipement manque de mémoire au cours d'une journée ?").

— De manière générale, les méthodes statistiques prennent une importance grandissante dans de nombreux domaines de l'informatique, et de ses applications au sens large. Les techniques d'*apprentissage* qui sont au coeur de nombreuses applications modernes, qu'il s'agisse de détecter un visage dans une image, de faire des suggestions d'achat au client d'un site de e-commerce, ou de battre les meilleurs joueurs humains au jeu de go – tous ces domaines ont été bouleversés au cours des dernières décennies par l'irruption de modèles probabilistes et de techniques statistiques.

Le présent cours n'a pas la prétention de vous amener à un niveau de maîtrise qui vous permettrait, sans compléments, de traiter tous ces exemples, mais plutôt de vous donner des bases les plus solides possibles dans ces domaines.

---

que la marge d'erreur.

Chapitre 2

Probabilités discrètes

2.1 Notion(s) de hasard

Qu'est-ce que le hasard ? Les définitions qu'on peut trouver dans les dictionnaires tournent autour de la notion d'un "événement incertain, imprévisible" ou d'une "cause imprévisible, sort, destin"...L'existence même d'un hasard intrinsèque n'est pas forcément l'objet d'un consensus ; qu'on pense par exemple à la célèbre citation d'Einstein "Dieu ne joue pas aux dés", exprimant l'hostilité du physicien à l'interprétation probabiliste de la physique quantique (et donc, en un sens, à l'idée que, au niveau fondamental, le monde serait régi par des phénomènes intrinsèquement aléatoires).

Il ne faut pas se tourner vers la théorie des probabilités pour y trouver une définition du hasard. Comme toute théorie mathématique, celle-ci part d'axiomes qui servent de "règles de base" (ici, ces règles sont censées définir ce que devrait satisfaire la notion intuitive de probabilité), et en tire des conclusions. Tout ce que dit, au final, la théorie, c'est quelque chose comme "tout système de hasard qui se conforme à nos axiomes (qui sont naturels) devrait satisfaire les conclusions obtenues".

L'histoire de la théorie des probabilités est vieille de plusieurs siècles. Des penseurs comme Blaise Pascal s'y sont attaqués, cherchant initialement à répondre à des questions sur les "jeux de hasard" (à base de dés, de cartes) qui faisaient fureur à leur époque [1]

L'objet final de la théorie des probabilités consiste à définir, lors de telle ou telle "expérience", et pour un "événement" qui désigne un ensemble de résultats possibles de l'expérience, une mesure chiffrée de la "plausibilité" de cet événement, qu'on appellera sa probabilité ; mesure qu'arbitrairement on prendra entre 0 et 1, avec l'idée qu'un événement "impossible" se verra attribuer la probabilité 0, et un événement "certain", la probabilité 1.

Une proposition souvent avancée pour "définir" la probabilité d'un événement E, consiste à prendre la proportion asymptotique du nombre de "succès" si on répète indéfiniment la même expérience (dans les mêmes conditions) un nombre infini de fois. Il s'agirait donc d'une prise de limite : on note le nombre de fois e_n que "E" se produit parmi les n premières expériences, et on prend comme définition de la probabilité de E, $p = \lim_{n\to\infty} e_n/n$.

Une telle "définition" rencontre immédiatement des difficultés de fond : est-il possible de montrer qu'une telle limite existe ? Et peut-on prétendre ainsi définir la probabilité d'un "évé-

1. Un exemple : vaut-il mieux parier qu'on obtiendra un 6 en lançant 4 fois un dé, ou qu'on obtiendra un 12 en lançant 24 fois une paire de dés ? Dans les deux cas, l'expérience devrait montrer que les chances sont proches d'une chance sur deux...

nement" qui ne peut se produire qu'une seule fois ? (Comment puis-je évaluer la probabilité que j'ai de mourir renversé par une voiture si je traverse la rue sans faire attention ? Même en me supposant un total dévouement à la science, je ne peux mourir renversé qu'une seule fois...)

2.2 Le vocabulaire des probabilités

2.2.1 Les axiomes des probabilités

Nous commençons ce chapitre, avant de présenter la théorie telle qu'elle est ordinairement décrite, par établir un "cahier des charges" de ce que devrait intuitivement satisfaire le calcul des probabilités.

On va partir d'une description d'une "situation", d'une "expérience" faisant potentiellement intervenir ce fameux "hasard" : quelque chose va se produire, mais on ne sait pas complètement quoi, et on est prêt à accepter que, peut-être, ce n'est pas complètement prévisible.

Notre objectif est d'évaluer la "plausibilité" de certaines possibilités, et d'attribuer à chacune de ces possibilité un nombre, qu'on appellera sa probabilité. En gros, on va évaluer les "chances" d'un résultat, ou d'un ensemble de résultats, possible, sur une échelle de 0 à 1 – 0 désignant l'impossible, 1 le certain...et plus le nombre est élevé, plus la possibilité semble forte.

On peut essayer de décrire notre échelle plus précisément : 0.5 représente "une chance sur deux" – disons que cela correspond à une situation sur laquelle on est raisonnablement prêt à parier à égalité de mise, sans préférence pour le côté sur lequel on parie : "si la pièce tombe côté pile, je te donne mon pain au chocolat ; si elle tombe côté face, tu me donnes ta chocolatine" [2]

Commençons, donc. On va se rapprocher légèrement du vocabulaire classique des probabilistes, et appeler "événement" une chose dont on s'apprête à calculer la probabilité (le mot est arbitraire, il est entré dans les usages).

On sait déjà que *si un événement est impossible, alors sa probabilité doit être de* 0. Par exemple, si je lance un dé classique à 6 faces, l'événement "le dé s'arrête sur une face qui indique 7" ne peut pas se produire : il n'y a pas de telle face.

De la même manière, on peut convenir que *si un événement est certain, alors sa probabilité doit être de* 1. Si je joue à pile ou face avec une pièce, et que j'exclus les possibilités "exotiques" comme "la pièce ne retombe pas", "la pièce disparaît avant de s'arrêter", ou "la pièce tombe en équilibre sur sa tranche", alors l'événement "la pièce donne pile ou face" devrait avoir probabilité 1.

On peut aussi dire que *si un événement A ne peut pas se produire sans que B se produise, alors la probabilité de A ne peut pas être plus petite que celle de B.* Par exemple, l'événement A "le dé donne 2", entraine forcément "le dé donne un résultat pair", parce que 2 est pair (il y a d'autres possibilités pour que le dé donne un résultat pair : 4 et 6 ; donc je suis plus prêt à parier que le résultat sera pair, que sur le fait que le résultat sera 2 – même si je soupçonne le dé d'être pipé).

2. Au-delà d'essayer de situer la ligne de démarcation d'une polémique nationale, cet exemple a ses limites : si j'ai très faim, je ne serai pas forcément prêt à parier mon unique pain au chocolat dans l'espoir d'en obtenir un deuxième ; mais ce serait une considération de modélisation économique : tenons-nous-en aux probabilités, c'est bien assez difficile comme ça.

Est-ce que c'est tout ? Il reste au moins une règle qui ne devrait pas trop causer de polémique : *si deux événements A et B ne peuvent pas se produire simultanément, alors l'événement "A se produit, ou B se produit", devrait avoir une probabilité qui soit la* ***somme*** *des probabilités respectives de A et de B.* Si par exemple j'estime la probabilité de l'événement "le dé donne 3" à 0.25, et celle de l'événement "le dé donne 6" à 0.12 (c'est un dé très pipé), alors je devrais estimer la probabilité de l'événement "le dé donne 3 ou 6" à 0.37, parce que le dé ne peut absolument pas donner **à la fois** 3 et 6.

Accolée à la règle sur les événements certains, cette règle nous en donne une autre : l'événement "contraire de A" ("A ne se produit pas") ne peut pas se produire simultanément avec A, et, de manière certaine, l'un des deux doit se produire. Donc, la somme des probabilités respectives de A et de "non A" doit être 1 : si le premier a pour probabilité p, le second a forcément pour probabilité $1 - p$.

Et la probabilité que *deux événements se produisent en même temps* ? On peut écrire une équation ? En toute généralité, on ne peut pas, ça dépend des événements. On peut juste dire que, si "A et B" se produit, alors A se produit, donc on peut prédire que la probabilité de "A et B" est inférieure ou égale à celle de A (et aussi à celle de B, pour le même type de raison), mais ça ne donne pas une règle de calcul.

Il y a une situation où on peut proposer une règle de calcul, c'est si on sait que, grosso modo, "les causes qui peuvent entrainer A ne peuvent aucunement influencer sur le fait que B se produise ou non" : on parle d'*indépendance* pour A et B, et dans ce cas la règle est que la probabilité de "A et B" est le **produit** des probabilités respectives de A et B. Par exemple, si j'ai un dé noir dont j'estime qu'il a une chance sur 10 (proba 0.1) de donner le résultat 3 (soit il est pipé, soit c'est un dé à 10 faces), et par ailleurs un dé blanc dont j'estime qu'il a une chance sur 2 (proba 0.5) de donner un résultat pair, alors je peux estimer que si je lance les deux dés, j'ai une chance sur 20 (probas $0.1 \times 0.5 = 0.05$) de voir le premier donner 3 et le second donner un nombre pair (les plus soupçonneux lanceront un dé, puis l'autre, après avoir éloigné le premier, des fois qu'ils s'influencent l'un l'autre).

On arrive à un ensemble assez denses de règles, qui nous permettent de faire une sorte de calcul logique, à base de "et", de "ou", et de "non". Récapitulons : si je note $p(A)$ la probabilité d'un événement A, alors je devrais avoir

1. $p(\text{impossible}) = 0$, $p(\text{certain}) = 1$.
2. Si "A et B" est impossible, alors $p(A \text{ ou } B) = p(A) + p(B)$.
3. $p(\text{non-}A) = 1 - p(A)$.
4. Si "A entraine B", alors $p(A) \leq p(B)$.
5. (Indépendance) Si A et B peuvent être considérés comme "indépendants", alors $p(A \text{ et } B) = p(A).p(B)$.

On ne s'en sort pas trop mal, on dirait... en fait, la règle 4 peut être vue comme une conséquence de la règle 2 (si A entraine B, alors B peut être décrit comme "A ou (B et non-A)"...). De plus, la règle 5 est trop imprécise pour les mathématiciens, qui n'aiment pas parler de causes ; ils vont plutôt la prendre comme *définition* de l'indépendance d'événements (avec comme conséquence, le fait que deux événements peuvent avoir plein de causes communes, et être indépendants ; nous verrons des exemples).

C'est tout ? Presque...et pas du tout, à la fois. La règle 2 est très bien, mais elle n'est pas assez forte pour faire certains calculs. Avec elle on peut passer de 2 événements à 3, à 4, à n'importe quel nombre fini ; mais on ne peut pas passer à un nombre *infini* d'événements.

Et parfois, on a besoin d'exprimer une condition comme un "ou" d'une infinité de conditions incompatibles et qui ont chacune leur petite probabilité ; l'exemple le plus simple serait celui d'une expérience qui pourrait potentiellement donner n'importe quel nombre entier naturel $(0, 1, \ldots)$, mais où on voudrait évaluer la probabilité de l'événement "l'entier donné est pair" : il y a une infinité de nombres pairs, mais il y a aussi une infinité de nombres non pairs (sinon on pourrait s'en sortir avec la règle 3).

Il se trouve que, mathématiquement, il y a une théorie qui fonctionne bien pour sommer des infinités (dénombrables) de nombres, surtout s'ils sont positifs. Et c'est pour cette raison que les axiomes classiques des probabilités incluent une règle un peu plus compliquée que notre règle 2 : au lieu d'être juste "additives", les probablités sont "σ-additives" ("sigma-additives"). En gros, il va falloir, parfois, prendre des limites.

2.2.2 Univers, loi de probabilités, événement

Les fondements axiomatiques de la théorie moderne des probabilités sont dûs au mathématicien russe Kolmogorov, au début du 20e siècle. L'astuce consiste à considérer un "univers" (un ensemble) des résultats possibles, sous la forme d'un ensemble abstrait Ω, et à associer à chaque partie E de Ω un nombre (appelé sa probabilité), en s'assurant que *collectivement* les probabilités des différentes parties satisfont à certaines règles, qui sont essentiellement celles qui ont été discutées dans ce qui précède.

Dans ce chapitre, les ensembles sont "discrets" : finis, ou infinis mais dénombrables.

Définition 2.1 ***(Univers de probabilités)*** *On appelle* univers de probabilités discret *un ensemble fini ou dénombrable Ω. Une probabilité $\mathbb{P}$ sur Ω est alors une application*

$$\begin{array}{rcl} \mathbb{P} : \mathit{Parties}(\Omega) & \to & [0,1] \\ A & \mapsto & \mathbb{P}(A) \end{array}$$

qui satisfait les propriétés suivantes :

- $\mathbb{P}(\Omega) = 1$
- $\mathbb{P}$ *est* σ-additive *: si $A_1, A_2, \ldots$ sont des parties de Ω qui sont* deux à deux disjointes *(c'est-à-dire $A_i \cap A_j = \emptyset$ dès que $i \neq j$), alors on a*

$$\mathbb{P}\left(\cup_i A_i\right) = \sum_i \mathbb{P}\left(A_i\right), \tag{2.1}$$

  formule qui doit être valable pour n'importe quel nombre fini (on parle d'additivité) ou infini dénombrable (σ-additivité) d'ensembles A_i.

La paire $(\Omega, \mathbb{P})$ est alors appelée espace probabilisé.

On dira également que $\mathbb{P}$ est une probabilité *sur Ω, ou* une loi de probabilités *sur Ω.*

Cette définition est suffisamment complexe pour qu'on prenne un peu le temps de la décortiquer.

Ici, la notation Parties(Ω) désigne l'ensemble de tous les sous-ensembles de Ω, y compris l'ensemble vide $\emptyset$ et l'ensemble complet Ω. Lorsque A et B sont des sous-ensembles de Ω, les notations $A \cup B$ ("A union B") et $A \cap B$ ("A inter B") désignent respectivement la réunion et l'intersection de A et B :

- $A \cup B$ contient tous les éléments de Ω qui sont présents dans A ou dans B (au sens inclusif : ceux qui sont et dans A et dans B sont bien dans $A \cup B$;

— $A \cap B$ contient tous les éléments de Ω qui sont à la fois dans A et dans B (et A et B sont *disjoints* si $A \cap B = \emptyset$, *i.e.* si aucun élément n'est à la fois dans A et dans B.

Ces notations sont étendues pour définir l'union ou l'intersection d'une famille potentiellement infinie de sous-ensembles $(A_i)_{i\in I}$ (ici c'est I qui peut être infini) :

— $\cup_{i\in I} A_i$ désigne l'union de tous les A_i : un élément x de Ω est dans l'union s'il est dans *au moins un* des A_i ;

— $\cap_{i\in I} A_i$ désigne l'intersection de tous les A_i : un élément x de Ω est dans l'intersection s'il est dans *chacun* des A_i.

Lorsque l'on a une famille *finie* $(A_i)_{1\leq i\leq n}$ de sous-ensembles (deux à deux disjoints), (2.1) s'interprète simplement, sous la forme

$$\mathbb{P}\left(A_i \cup A_2 \cdots \cup A_n\right) = \sum_{i=1}^{n} \mathbb{P}(A_i); \tag{2.2}$$

dans le cas où il s'agit d'une famille *infinie dénombrable* $(A_i)_{i\geq 1}$ (i parcourt ici l'ensemble des entiers strictement positifs), il faut l'interpréter comme une *limite* :

$$\mathbb{P}\left(\cup_{i\geq 1} A_i\right) = \lim_{n\to\infty} \left(\sum_{i=1}^{n} \mathbb{P}(A_i) \right). \tag{2.3}$$

On peut naturellement se demander si la complexité apportée par ce besoin de traiter des sommes infinies (séries) est bien nécessaire ; la réponse est malheureusement oui, comme on le verra, dès lors que Ω est un ensemble infini.

Avant d'explorer plus en détails cette notion d'espace probabilisé, vérifions qu'elle permet de rendre compte de situations intuitivement raisonnables. L'idée est que Ω représente l'ensemble de *tous les résultats possibles d'une expérience* (ou, dans le cas d'une suite de plusieurs expériences, de *toutes les séquences de résultats possibles*), et, pour une sélection quelconque de certains résultats possibles (un sous-ensemble $E \subset \Omega$), $\mathbb{P}(E)$ représente le "poids", la "vraisemblance" (on dira la probabilité) de l'ensemble des éléments de E.

Exemple 2.2
— *Lancer d'une pièce légèrement biaisée : on peut prendre* $\Omega = \{pile, face\}$, *et par exemple* $\mathbb{P}(\{pile\}) = 0.51$ *et* $\mathbb{P}(\{face\}) = 0.49$.

— *Lancer d'un dé équilibré : on prend* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *et, pour tout* $i \in \Omega$, $\mathbb{P}(\{i\}) = 1/6$.

— *Tirage à pile ou face (avec une pièce équilibrée) jusqu'à obtenir face pour la première fois : on peut prendre* $\Omega = \mathbb{N}^* = \{1, 2, \ldots\}$, *et* $\mathbb{P}(\{i\}) = 1/2^i$ *pour tout* $i \in \Omega$ *(cet exemple est un peu plus complexe ; comme* Ω *est infini, il y a besoin d'un petit calcul pour vérifier, en utilisant (2.1), que l'on a bien* $\mathbb{P}(\Omega) = 1$*).*

Définition 2.3 *Les éléments de* Ω *sont appelés* événements élementaires *; ceux de Parties*(Ω) *sont appelés* événements.

Parmi les événements, il y en a qui sont composés d'un seul événement élémentaire, *i.e.* les singletons. Il est tentant de simplifier les notations, et d'écrire, si $x \in \Omega$ (x est un "événement élémentaire"), $\mathbb{P}(x)$ pour $\mathbb{P}(\{x\})$. Cette notation est acceptable, du moment qu'on garde en tête qu'il s'agit d'un raccourci d'écriture.

Il est maintenant temps de vérifier que la théorie qu'on est en train d'esquisser, n'est pas en contradiction avec l'idée intuitive (ou même avec une autre définition qu'on a pu voir dans sa prime jeunesse !) de ce qu'est une probabilité.

Proposition 2.4 *Soit $(\Omega, \mathbb{P})$ un espace probabilisé. On a alors*

1. $\mathbb{P}(\emptyset) = 0$.
2. *Si A et B sont deux parties de Ω, et que l'on a $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.*
3. *Si A est une partie de Ω, et si A^c désigne le* complémentaire *de A dans Ω (c'est-à-dire, $A^c = \Omega \backslash A$; A^c est l'ensemble des éléments de Ω qui ne sont pas dans A), alors on a*

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

4. *Si A et B sont deux parties quelconques de Ω, alors on a*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Preuve:

1. Ω et $\emptyset$ sont deux parties de Ω, et ils sont bien disjoints ; en appliquant (2.1), on a donc $\mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset)$, qui implique $\mathbb{P}(\emptyset) = 0$.
2. Par définition, A et A^c sont disjoints, et leur union est Ω ; en appliquant aussi (2.1), on a donc $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$ qui donne la formule voulue.
3. Supposons $A \subset B$, et posons $C = B \backslash A$ (C est formé des éléments de B qui ne sont pas dans A). A et C sont ainsi deux parties disjointes de Ω, et $A \cup C = B$. On a donc $\mathbb{P}(A) + \mathbb{P}(C) = \mathbb{P}(B)$; comme $\mathbb{P}(C) \geq 0$, on en déduit $\mathbb{P}(B) \geq \mathbb{P}(A)$.
4. (Faire un dessin peut aider ici) On peut poser $C = A \cap B$, $D = A \backslash C$, et $E = B \backslash C$. Les trois parties C, D et E sont deux à deux disjointes, et on a $A \cup B = C \cup D \cup E$. En appliquant la formule précédente, on obtient successivement

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(C) + \mathbb{P}(D) \\ \mathbb{P}(B) &= \mathbb{P}(C) + \mathbb{P}(E) \\ \mathbb{P}(A \cup B) &= \mathbb{P}(C) + \mathbb{P}(D) + \mathbb{P}(E) \end{aligned}$$

   Si l'on somme les deux premières équations ci-dessus, il faut retrancher $\mathbb{P}(C)$ pour que le second membre devienne celui de la troisième équation ; on obtient alors la formule proposée pour $\mathbb{P}(A \cup B)$.

□

Une autre propriété rassurante est la suivante : dans un univers fini ou dénombrable, la probabilité est entièrement définie si on connait sa valeurs sur les *singletons*.

Proposition 2.5 *Soit Ω un ensemble fini ou dénombrable. Supposons que l'on ait une "fonction de poids" sur Ω : une fonction $p : \Omega \to \mathbb{R}$ telle que pour tout $x \in \Omega$, $p(x) \geq 0$, et par ailleurs $\sum_{x \in \Omega} p(x) = 1$.*

Alors il existe une et une seule loi de probabilité $\mathbb{P}$ sur Ω telle que l'on ait, pour tout $x \in \Omega$, $\mathbb{P}(\{x\}) = p(x)$.

Preuve: Il y a deux choses à prouver : d'une part, qu'il existe bien une telle loi de probabilités ; d'autre part, qu'elle est unique. On ne va pas donner la preuve complète, qui fait un peu trop appel à la manipulation de séries.

Commençons par l'existence : nous allons définir $\mathbb{P}$, et vérifier qu'il s'agit bien d'une loi de probabilités. L'énoncé nous a déjà défini $\mathbb{P}$ pour les parties qui sont des singletons ; partant de là, comme nous voulons satisfaire (2.1), il est naturel de poser, pour tout $A \subset \Omega$,

$$\mathbb{P}(A) = \sum_{x \in A} p(x).$$

On obtient une définition de $\mathbb{P}(A)$ pour tout $A \subset \Omega$, qui donne bien $\mathbb{P}(\{x\}) = p(x)$; il faudrait vérifier qu'il s'agit bien d'une loi de probabilités, ce que nous ne ferons pas (caché derrière : comme on manipule des sommes de réels positifs, "on a le droit" de faire les manipulations de sommes dont on a besoin pour vérifier la σ-additivité de $\mathbb{P}$).

Pour l'unicité, c'est beaucoup plus facile : pour n'importe quelle loi de probabilité $\mathbb{P}'$ qui satisfait l'hypothèse, et pour n'importe quelle partie $A \subset \Omega$, on peut écrire A comme union (finie ou dénombrable) de singletons disjoints : $A = \cup_{x \in A}\{x\}$; et donc, en appliquant la σ-additivité, on a $\mathbb{P}'(A) = \sum_{x \in A} \mathbb{P}'(\{x\}) = \sum_{x \in A} p(x) = \mathbb{P}(A)$. □

Le lecteur particulièrement attentif aura remarqué que nous avons déjà utilisé cette proposition sans le dire : dans l'exemple 2.2, à chaque fois nous avons seulement décrit la probabilité par sa valeur sur les singletons.

Revenons à la question de la σ-additivité, et posons-nous la question naturelle : pourrait-on, dans la définition, ne demander que l'additivité (équation (2.2)) au lieu de la σ-additivité (équation (2.1) ? Ce serait bien plus simple...

Malheureusement, si on fait cela, les choses ne se passent pas aussi bien que prévu, et en particulier, il n'est plus vrai qu'une "loi de probabilités" est définie par ses valeurs sur les singletons.

Exercice 2.1 *Prenons $\Omega = \mathbb{N}$, et définissons deux "pseudo-lois" de probabilité sur Ω, de la manière suivante : pour tout $A \subset \Omega$,*

$$\mathbb{P}(A) = \begin{cases} 0 & \text{si } A \text{ est fini} \\ 1 & \text{si } A \text{ est infini} \end{cases} \qquad \mathbb{P}'(A) = \begin{cases} 0 & \text{si } A^c \text{ est infini} \\ 1 & \text{si } A^c \text{ est fini} \end{cases}$$

Montrer que $\mathbb{P}$ et $\mathbb{P}'$ sont bien additives (équation (2.2), qu'elles satisfont la proposition 2.4, et que pour tout $n \in \Omega$, on a bien $\mathbb{P}(\{n\}) = \mathbb{P}'(\{n\})$. Donner un exemple de partie $A \subset \Omega$ pour laquelle on a $\mathbb{P}(A) \neq \mathbb{P}'(A)$.

2.2.3 Le cas de l'équiprobabilité

Lorsque l'univers de probabilités est un ensemble *fini*, on peut définir une loi de probabilité par simple comptage, et on obtient ce que l'on appelle la *loi uniforme*.

On suppose ici que les ensembles considérés sont finis, et on note $\#A$ le cardinal d'un ensemble A (son nombre d'éléments).

Définition 2.6 *Soit Ω un univers fini. On appelle* loi uniforme*, ou* équiprobablité*, sur Ω, la loi définie de la manière suivante : pour tout $A \subset \Omega$, on pose*

$$¶(A) = \frac{\#A}{\#\Omega}. \tag{2.4}$$

De manière équivalente, on peut définir cette loi uniforme par sa valeurs sur les singletons : pour tout $x \in \Omega$,

$$\P(x) = \frac{1}{\#\Omega}. \tag{2.5}$$

La loi uniforme est donc une loi qui accorde à chaque événement élémentaire la même probabilité. Il n'est pas difficile de vérifier que c'est la seule qui ait cette propriété : prendre une autre valeur que $1/\#\Omega$ pour la probabilité des événements élémentaires ne donnerait pas $\P(\Omega) = 1$.

Lorsque l'on prend un modèle d'équiprobabilité, on a donc la formule qui peut sembler familière pour la probabillité de n'importe quel événement : "nombre de cas favorables, divisé par le nombre total de cas". Cette formule n'est bien sûr valide que dans le cas de l'équiprobabilité ; il est important de bien retenir que ce n'est qu'un cas particulier.

2.2.4 Description des événements (et abus de notation)

On a déjà commis, dans les pages qui précèdent, un abus de notation, consistant à noter, lorsque x est un événement élémentaire (un élément de l'univers), $\P(x)$ alors qu'il faudrait normalement noter $\P(\{x\})$.

Il y a un autre abus de notation que les probabilistes s'autorisent régulièrement, et nous ferons de même : c'est celui qui consiste à noter un événement (qui, formellement, est une partie de l'univers, c'est-à-dire un ensemble d'événements élémentaires) par la description de la condition qui le définit.

Comme c'est un peu abstrait, prenons un exemple – celui du lancer d'un dé équilibré. On le définit en prenant $\Omega = \{1, 2, 3, 4, 5, 6\}$, avec la loi uniforme ; chaque événement élémentaire correspond à l'un des résultats possibles.

Dans ce contexte, l'événement "le résultat est pair" correspond à $\{2, 4, 6\}$ (l'ensemble des événements élémentaires qui sont des nombres pairs), et "le résultat est multiple de 3", à $\{3, 6\}$.

La loi de probabilité choisie étant uniforme (c'est l'hypothèse que le dé est équilibré), on calcule simplement les probabilités de ces deux événements : le premier a probabilité $1/2$ $(3/6)$, et le second, $1/3$ $(2/6)$. On s'autorisera ainsi à écrire

$$\P(\text{le résultat est pair}) = \frac{1}{2}$$

ou

$$\P(\text{le résultat est impair}) = \frac{1}{3}$$

On peut également s'autoriser à exprimer les événements en formant des conditions au moyen de connecteurs logiques comme "et" et "ou" : dans notre exemple, l'événement "le résultat est pair ou multiple de 3" est donc $\{2, 3, 4, 6\}$ et a pour probabilité $2/3$, et l'événement "le résultat est pair et multiple de 3" se réduit à $\{6\}$.

Le connecteur "et" correspond ainsi à l'opération ensembliste d'intersection, et le "ou", à l'union : les éléments de Ω qui satisfont à la condition A **et** à la condition B, sont ceux qui sont **à la fois** parmi ceux qui satisfont à la condition A et parmi ceux qui satisfont à la condition B.

Ne pas confondre avec la recette pour former une union ou une intersection, qui inverse la correpondance : pour former l'ensemble $A \cup B$ (corrspondant à un "ou"), on prend les éléments de A **et** les éléments de B...

En particulier, quand on cherche à former une condition un peu plus complexe qui implique les deux types de connecteurs, il est indispensable d'utiliser des parenthèses.

2.2.5 Probabilité conditionnelle et indépendance

La notion de probabilité conditionnelle est fondamentale, et souvent mal comprise. Elle correspond à la situation suivante : une expérience aléatoire est effectuée, et on apprend non pas le résultat, mais une information partielle sur ce résultat – pour utiliser le vocabulaire introduit jusqu'ici, on n'apprend pas quel événement élémentaire s'est produit, mais seulement qu'un certain événement B (non élémentaire) s'est produit.

Dans une telle situation, cette information modifie notre estimation des probabilités : pour un $x \in \Omega$ qui ne fait pas partie de B ($x \notin B$), nous savons que ce n'est pas lui qui a été choisi : notre estimation de sa probabilité passe donc à 0. En revanche, si y et z sont tous les deux dans B, leur probabilité reste positive, et l'information "B se produit" ne devrait pas changer le *ratio* de leurs probabilités. Si, par exemple, y était deux fois plus probable que z, il reste deux fois plus probable que z une fois qu'on apprend que B (qui contient à la fois y et z) se produit.

Ces considérations amènent naturellement à la définition suivante.

Définition 2.7 *Soit $(\Omega, \mathbb{P})$ un espace probabilisé, et B un événement tel que $\P(B) > 0$.*

Alors pour tout événement A, la probabilité de A sachant B, notée $\mathbb{P}(A|B)$ ou $\mathbb{P}_B(A)$, est définie par

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

De manière équivalente, cette probabilité peut être décrite par sa valeur sur les singletons : pour tout $x \in \Omega$,

$$\mathbb{P}(x|B) = \begin{cases} \frac{\mathbb{P}(x)}{\mathbb{P}(B)} & \text{si } x \in B \\ 0 & \text{si } x \notin B \end{cases}$$

On vérifie aisément, à partir de la définition, les faits suivants : d'une part, $\mathbb{P}_B$ ainsi définie est bien une loi de probabilités sur Ω ; d'autre part, si la probabilité "de départ" $\mathbb{P}$ est la loi uniforme sur Ω, alors la loi $\mathbb{P}_B$ est assimilable à la loi uniforme sur B.