ministère
éducation
nationale

ÉduSCOL
[dgesco]

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Mathématiques

Collège

- Ressources pour les classes de 6^{e}, 5^{e}, 4^{e}, et 3^{e} du collège -

- Probabilités au collège -

Mars 2008

Probabilités

« Pour comprendre l'actualité, une formation à la statistique est aujourd'hui indispensable ; c'est une formation qui développe des capacités d'analyse et de synthèse et exerce le regard critique. Le langage élémentaire de la statistique (avec ses mots tels que moyenne, dispersion, estimation, fourchette de sondage, différence significative, corrections saisonnières, espérance de vie, risque, etc.) est, dans tous les pays, nécessaire à la participation aux débats publics : il convient donc d'apprendre ce langage, ses règles, sa syntaxe, sa sémantique ; l'enseignement de la statistique étant, par nature, associé à celui des probabilités, il s'agit en fait d'une " formation à l'aléatoire ". »[1]. Le rapport de la commission de réflexion sur l'enseignement des mathématiques, d'où la citation précédente est extraite, évoque dans les termes suivants l'enseignement au collège et au lycée : « L'objectif d'une initiation aux probabilités et à la statistique aux niveaux collège et lycée est d'enrichir le langage, de repérer des questions de nature statistique, de définir des concepts qui fonderont un mode de pensée pertinent, rassurant, remarquablement efficace. Les modes usuels de représentation graphique (histogrammes, diagrammes en bâtons notamment), c'est-à-dire les éléments de base du langage graphique de la statistique sont aujourd'hui enseignés en collège et une introduction à l'aléatoire, appuyée sur le calcul de probabilités et la simulation, est proposée dans les nouveaux programmes de lycée ». La mise en place du socle commun modifie cette répartition, en demandant que les élèves, à la fin de la scolarité obligatoire, connaissent « les notions de chance ou de probabilité ». Alors qu'un enseignement des probabilités a depuis longtemps trouvé sa légitimité au niveau du lycée, un tel enseignement est une nouveauté en France au niveau du collège, contrairement à la situation existant dans de nombreux pays voisins (Allemagne, Espagne, …). Le programme de troisième comporte la rubrique reproduite ci-dessous :

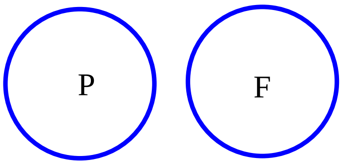
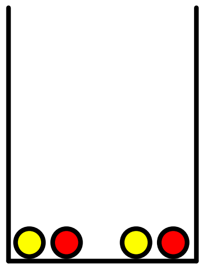
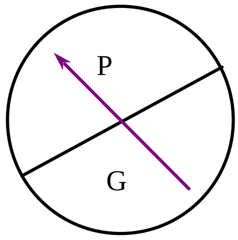
1.4. Notion de probabilité [Thèmes de convergence]	- Comprendre et utiliser des notions élémentaires de probabilité. - Calculer des probabilités dans des contextes familiers.	La notion de probabilité est abordée à partir de situations familières (pièces de monnaie, dés, roues de loteries, urnes). Certaines de ces situations permettent de rencontrer des cas pour lesquels les probabilités ne sont pas définies à partir de considérations intuitives de symétrie ou de comparaison mais sont approximativement évaluées par les fréquences observées expérimentalement (approche fréquentiste des probabilités). La notion de probabilité est utilisée pour traiter des situations de la vie courante pouvant être modélisées simplement à partir des situations précédentes. Les situations étudiées concernent les expériences aléatoires à une ou *à deux épreuves*.	. Dans le cadre du socle, aucune compétence n'est exigible dans le cas des expériences à deux épreuves.

Ce document **explicite les choix faits dans le programme de troisième**, en précisant dans les paragraphes 1 et 2 les contextes qui seront privilégiés dans les premières situations d'enseignement, dans les paragraphes 3 et 4 les moyens de représentation et de traitement.
Le paragraphe 5 **traite des expériences aléatoires à deux épreuves**, **sur lesquelles aucune compétence n'est exigible dans le cadre du socle commun**.
L'annexe 1 donne les diverses interprétations de la notion de chance (ou probabilité) et l'annexe 2 donne quelques éléments sur l'émergence de la notion de probabilité. Dans ces annexes, certains développements ou calculs, justifiant les résultats importants, font référence à la théorie enseignée dans l'enseignement supérieur.

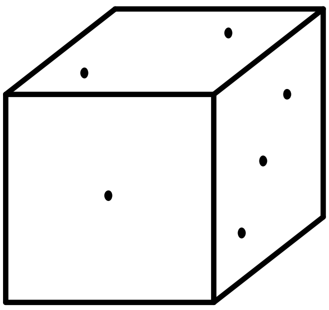
[1] Voir en bibliographie l'ouvrage [1], pages 52 et 53.

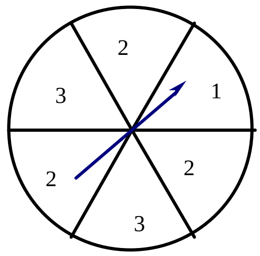
1. Probabilités définies à partir de considérations de symétrie ou de comparaison

Dans chacune des situations ci-dessous, deux issues (ou résultats) sont possibles, et on a 1 chance sur 2 de tirer “Pile”, de tirer une boule rouge, ou de tomber sur la région P, ce que l’on traduit en disant que la probabilité de chacune d’elles est égale à 1/2[2].

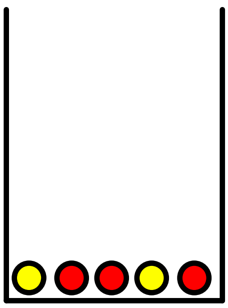
Lancer d’une pièce “équilibrée”	Tirage d’une boule dans une urne	Tirage avec une roue de loterie
		

D’autres situations classiques permettent d’obtenir d’autres valeurs pour les probabilités des différentes issues (ou événements) :



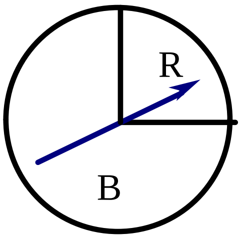


Les dispositifs précédents peuvent être adaptés pour mettre en évidence des événements n’ayant pas la même probabilité.



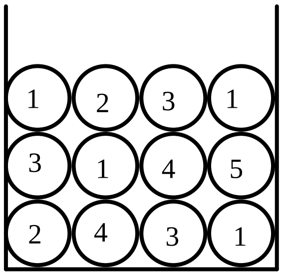
Dans le tirage au hasard d’une boule dans l’urne,
- on a 3 chances sur 5 d’obtenir une boule rouge.
- la probabilité d’obtenir une boule jaune est 2/5. Il y a 40% de chance d’obtenir une boule jaune.

[2] Les justifications solliciteront l’une quelconque des interprétations de la probabilité (cf. annexe 1) : interprétation fréquentiste dans sa variante « propension » ; mais certains élèves feront certainement appel à l’interprétation épistémique, dans sa variante personnelle ou interpersonnelle ; la variante logique conduisant à faire appel au principe d’indifférence (ou de raison insuffisante).



En lançant cette roue de loterie « équilibrée », la probabilité de tomber sur la région R est 1/4, …

Par tirage dans une urne ayant la composition suivante :

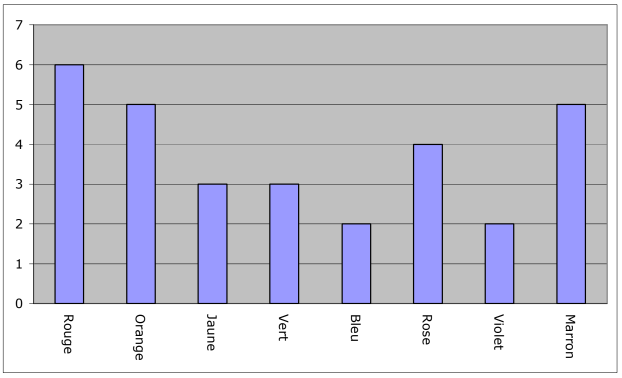


où chacune des boules a la même probabilité d'être tirée, les « résultats » 1, 2, 3, 4 et 5 ont respectivement comme probabilités : 1/3, 1/6, 1/4, 1/6 et 1/12. L'exploitation de tels exemples peut déboucher sur la mise en place de la formule de Laplace : la probabilité d'un résultat est égale au quotient du nombre d'issues favorables (issues dans lesquelles on obtient le résultat) par le nombre total d'issues possibles lors du tirage. Par exemple, quatre issues sont favorables au résultat « 1 », sur un total de 12 issues possibles.

On peut en déduire que la probabilité de l'événement « ne pas tirer une boule portant le numéro 1 » est égale à 8/12 ou 2/3, et que celle d'obtenir un résultat pair est égale à 1/3, premier contact avec la recherche de la probabilité d'un événement contraire ou de la probabilité d'obtenir l'un ou l'autre de deux événements.

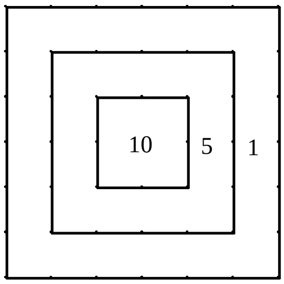
L'exercice suivant, tiré de l'évaluation PISA, fait appel à ce mode de calcul d'une probabilité :

La mère de Kevin lui permet de prendre un bonbon dans un sachet opaque. Kevin ne voit donc pas les bonbons. Le nombre de bonbons de chaque couleur contenus dans le sachet est illustré par le graphique suivant :



Quelle est la probabilité que Kevin prenne un bonbon rouge ?
A 10%
B 20%
C 25%
D 50%

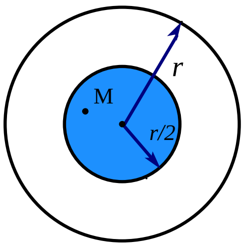
Des exemples de calculs de probabilités « géométriques » tels que ceux qui suivent permettent également de calculer les probabilités des événements du type « atteindre une région précise de la cible » ou « obtenir un écho radar dans une zone précise de l'écran de contrôle » :



On imagine qu'un tireur tire parfaitement au hasard sur la cible ci-contre, sans jamais la rater (!). Tous les carrés sont concentriques et leurs côtés ont pour mesure a, $2a$ et $3a$.
Quelles sont les probabilités pour qu'il gagne 10 points, 5 points, 1 point ?
La probabilité relative à une région est proportionnelle à son aire : c'est le rapport de son aire à celle de la cible.

Réponse : 1/9, 1/3 ou 3/9, 5/9.

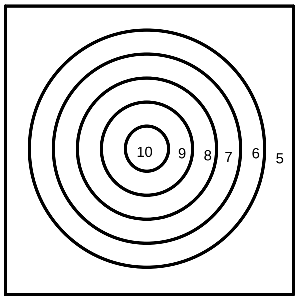
La recherche de la probabilité de tirer dans une région portant un numéro supérieur ou égal à 5 permet de mettre en place que l'on peut additionner les probabilités d'événements incompatibles ou qu'il est parfois plus facile de calculer d'abord la probabilité de l'événement contraire.



Sur l'écran circulaire de rayon r d'un radar, on suppose que le point M représentant un avion se projette au hasard sur l'écran.
Quelle est la probabilité pour qu'il apparaisse dans la zone colorée, disque de rayon $r/2$?

Réponse : 1/4.

On peut étudier des exemples plus compliqués du même type que le premier :



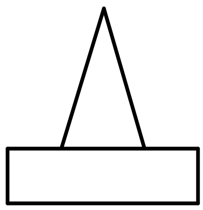
Un tireur tire parfaitement au hasard sur cette nouvelle cible, sans jamais la rater. Tous les cercles sont concentriques, leurs rayons sont r, $2r$, $3r$, $4r$, $5r$ et le carré a un côté de longueur $12r$.
Quelles sont les probabilités pour qu'un point d'impact appartienne à chacune des régions 10, 9, …, 5 ?

Réponse : 0,022 ; 0,065 ; 0,109 ; 0,153 ; 0,196 ; 0,455.
(Valeurs approchées).

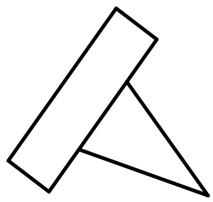
2. Approche fréquentiste de la probabilité

Les situations précédentes ne sont guère pertinentes pour aborder l'interprétation fréquentiste de la probabilité comme « fréquence limite ». Or cette interprétation est très importante pour les applications des probabilités dans des situations de la vie courante. Elle permet en outre de donner une justification des calculs de probabilités dans des expériences à deux épreuves, traités au paragraphe 5.

L'approche fréquentiste exige que des fréquences soient observées expérimentalement ; le lancer d'une punaise (pouvant tomber suivant la position « 1 » ou la position « 0 » ci-dessous) a longtemps servi d'exemple dans les pays anglo-saxons[3] :



1

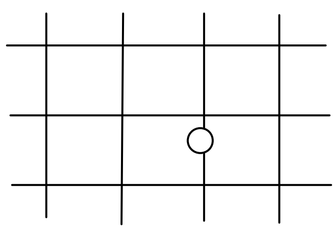


0

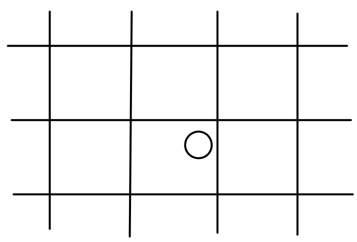
Pour un petit nombre de lancers successifs, la suite des résultats semble ne suivre aucune loi. Mais, en lançant un grand nombre de fois la punaise, la suite de résultats « 1 » et « 0 » laisse apparaître une régularité dans la fréquence de chacune des deux issues. Ainsi, les fréquences observées du résultat « 1 » en fonction du nombre de lancers connaissent au début une forte variabilité qui tend à se réduire avec le nombre de lancers. L'intérêt du lancer de punaise réside dans le fait que seule l'expérimentation permet de proposer une probabilité au résultat « 1 ».

Il est important, dans un premier temps, que les élèves puissent constater matériellement ce phénomène de stabilisation des fréquences. Toutefois, pour éviter les lancers de punaises (!), on peut proposer la situation du jeu du « Franc Carreau »[4], en cherchant à déterminer approximativement la probabilité de gagner.

Le jeu de « Franc Carreau » consiste à prendre une pièce de monnaie (de 1 cm de rayon, par exemple), et à la lancer sur un carrelage dont les carreaux sont des carrés (de 10 cm de côté, par exemple). On fait « Franc Carreau » quand la pièce tombe sur une seule case, dont elle peut toucher les bords, mais sans empiéter sur une autre case. Dans ce cas, on gagne un euro ; sinon, on perd un euro.



Le joueur perd 1€



« Franc Carreau » : le joueur gagne 1 €

Le joueur a-t-il davantage de chance de gagner que de perdre ?

L'idée d'entreprendre une série de lancers et de s'intéresser à la fréquence de succès est alors assez naturelle, et s'appuie sur la connaissance naïve de la loi des grands nombres

[3] Voir [8] en bibliographie.
[4] Par exemple, on peut remplacer le carrelage du sol par un quadrillage sur papier.

évoquée par Bernoulli : « plus on fait d'observations, moins on risque de s'écarter de notre but »[5]. Pour augmenter le nombre de lancers, on peut mettre en commun les résultats obtenus par des groupes d'élèves, puis par l'ensemble de la classe[6].

Cette situation présente l'avantage que l'on peut déterminer cette probabilité à l'aide de considérations géométriques, sans que cette valeur soit connue au départ (comme c'est le cas avec le jeu de pile ou face avec une pièce « équilibrée », ou le jeu du lancer d'un dé cubique « non truqué »). On peut conduire cette justification avec les élèves en leur posant la question : « Dans quelle partie du carré doit se trouver le centre de la pièce pour que le joueur puisse faire " Franc Carreau " ? » La réponse (carré de côté 8 cm, dans l'exemple) permet de calculer la probabilité à l'aide du rapport des aires des deux carrés.

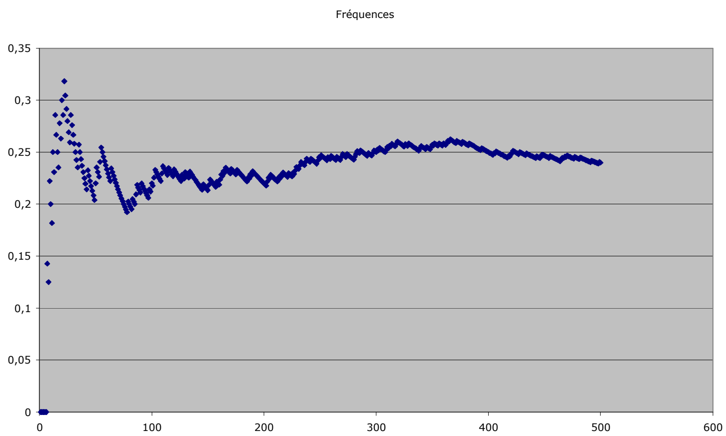
Dans un second temps, pour disposer facilement d'un grand nombre d'épreuves et interpréter graphiquement les résultats, on peut faire usage d'une simulation sur un tableur. Par exemple, la situation suivante :

> Sur un segment S, on prend au hasard deux points A et B. On considère l'événement « La longueur du segment [AB] est strictement supérieure à la moitié de celle du segment S ». Quelle est la probabilité de cet événement ?

peut être simulée à l'aide d'un tableur de la manière suivante. En prenant la longueur de S comme unité, A et B peuvent être déterminés par leurs abscisses qui sont des nombres compris entre 0 et 1, nombres que l'on peut obtenir à l'aide de la fonction ALEA().

	A	B	C	D	E	F
1	Expé-rience N°	Abscisse de A	Abscisse de B	Distance AB	AB > 0,5	Fréquences
2	1	=ALEA()	=ALEA()	=MAX(B2;C2)-MIN(B2;C2)	=SI(D2>0,5;1;0)	=E2/A2
3	=A2+1	=ALEA()	=ALEA()	=MAX(B3;C3)-MIN(B3;C3)	=SI(D3>0,5;1;0)	=SOMME(E$2:E3)/A3
4	=A3+1	=ALEA()	=ALEA()	=MAX(B4;C4)-MIN(B4;C4)	=SI(D4>0,5;1;0)	=SOMME(E$2:E4)/A4

En recopiant la dernière ligne jusqu'à ce qu'on obtienne par exemple 500 expériences, et en utilisant le calcul sur ordre[7], on obtient ainsi les fréquences relatives de l'événement dans autant de séries de 500 expériences qu'on le souhaite. En exploitant les ressources graphiques du tableur, on peut visualiser l'évolution des fréquences au fur et à mesure de l'augmentation du nombre d'expériences.

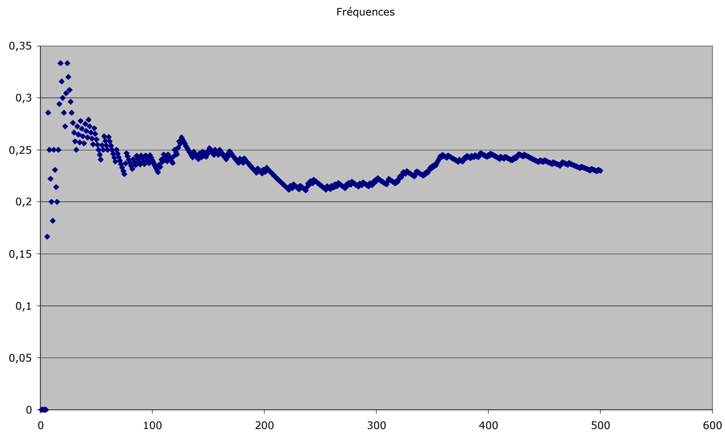


[5] Voir le paragraphe 2.1 de l'annexe 2.

[6] Le jeu du Franc Carreau est évoqué dans [9] et [10].

[7] Il convient pour cela de désactiver le mode de calcul automatique du tableur, et de passer dans un mode qui selon les logiciels s'appelle "calcul sur ordre" ou "recalculer", et s'obtient au clavier à l'aide de raccourcis utilisant la touche F9, selon des combinaisons dépendant du système d'exploitation.

Une telle représentation met en évidence que les fréquences obtenues pour un faible nombre d'expériences évoluent de manière assez chaotique, mais qu'à la longue, elles tendent à se stabiliser autour de 0,25, même si le phénomène n'est pas toujours aussi visible, comme le montre les résultats au cours d'une autre série de 500 expériences.



Ce dernier résultat n'est guère étonnant car l'intervalle de fluctuation au niveau 0,95 pour n égal à 500 est approximativement [20,5% , 29,5%] (Voir le paragraphe 2.2 de l'annexe 2).
L'emploi d'un tableur dans de telles simulations permet de donner une formulation naïve de la loi des grands nombres, que les élèves admettent volontiers : lorsque le nombre d'expériences augmente, la fréquence empirique se rapproche de la probabilité 0,25. D'une manière plus précise, on peut dire qu'il est d'autant plus probable que l'écart entre les deux soit proche de 0 que le nombre d'expériences n est grand.

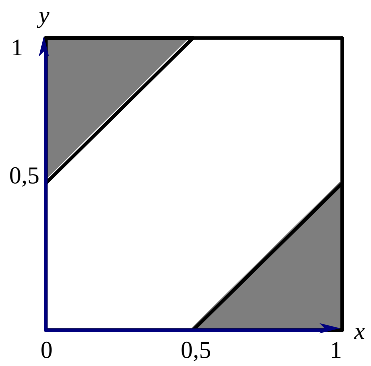
Cet exemple[8] présente le même avantage que le jeu du Franc Carreau : on peut justifier que 0,25 est bien la valeur de la probabilité, en ramenant la question à un problème de « probabilités géométriques ». Toutefois, la justification complète présentée ci-dessous est moins facile d'accès aux élèves que dans le cas du jeu du Franc Carreau.

Justification géométrique

On prend la longueur du segment comme unité, et on repère tout point du segment par son abscisse, nombre compris entre 0 et 1.
En désignant par x et y les abscisses respectives de A et B :

- si $y > x$, AB > 0,5 est équivalent à $y - x > 0{,}5$, c'est-à-dire $y > x + 0{,}5$.
- si $y \leq x$, AB > 0,5 est équivalent à $x - y > 0{,}5$, c'est-à-dire $x - 0{,}5 > y$, ou encore $y < x - 0{,}5$.

Représentons chaque segment [AB] par le point du plan dont les coordonnées dans un repère orthonormal sont (x, y).
Les segments [AB] dont la mesure est strictement supérieure à 0,5 sont ceux qui sont représentés par un point appartenant à la région grisée ci-dessous :



La probabilité recherchée est donc le rapport de l'aire grisée à celle du carré unité.

[8] Que l'on trouvera, entre autres, dans [11], pages 57 à 59.

Pour éviter les difficultés liées au régionnement qui précède, on peut se contenter d'une illustration avec un tableur, en faisant afficher sur un graphique les seuls points qui satisfont la condition : ils se répartissent dans la zone grisée ci-dessus.

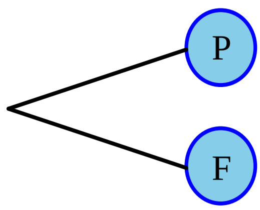
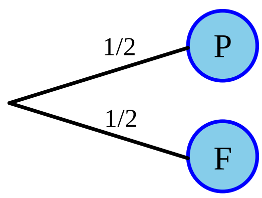
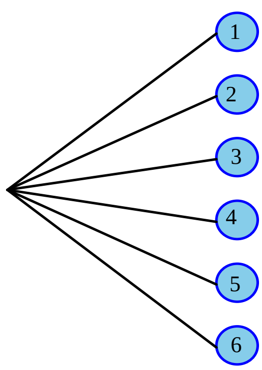
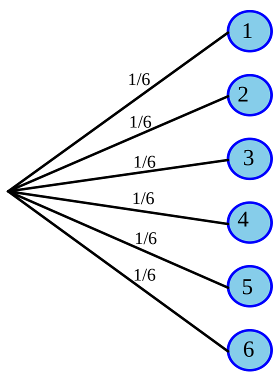
L'exercice suivant, tiré de l'évaluation PISA, sollicite la connaissance de l'interprétation fréquentiste de la probabilité.

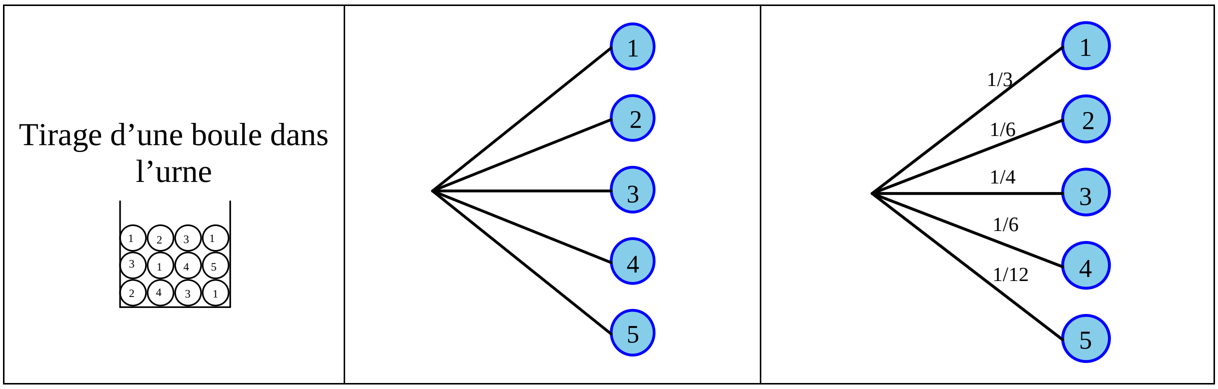
> Le bulletin météorologique du jour prévoit que, de 12 à 18 heures, les probabilités de pluie sont de 30 %.
>
> Laquelle des affirmations suivantes est la meilleure interprétation de ce bulletin ?
>
> A - Il va pleuvoir sur 30 % de la zone concernée par les prévisions.
> B - Il pleuvra pendant 30 % des 6 heures (un total de 108 minutes).
> C - Dans cette zone, 30 personnes sur 100 auront de la pluie.
> D - Si la même prévision était faite pour 100 jours, il pleuvrait à peu près 30 jours sur 100.
> E - La quantité de pluie tombée sera 30 % de celle tombée lors d'une forte pluie (mesurée en termes de précipitations par unité de temps).

La réponse attendue est la réponse D.

3. Moyens de représentation et de traitement

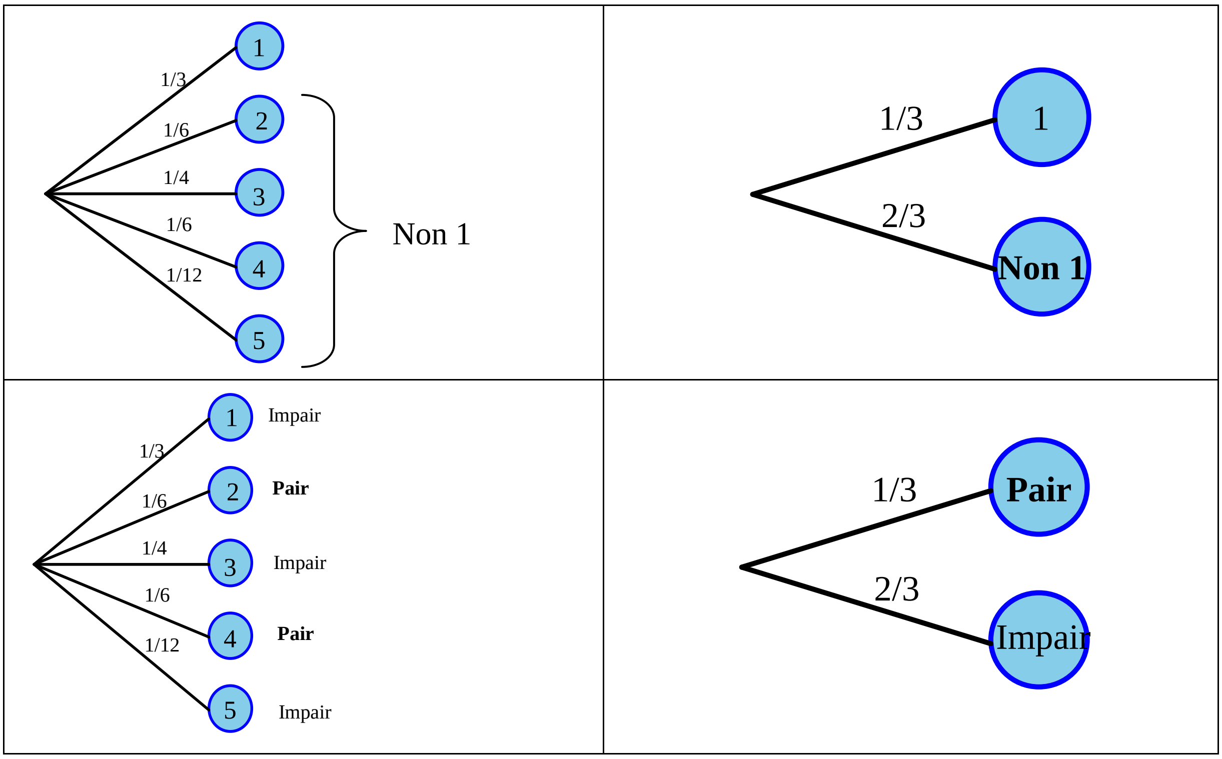
Les exemples d'expériences à une seule épreuve, évoqués ci-dessus, sont mis à profit pour mettre en place un moyen de représentation et de traitement qui sera réutilisé dans les expériences à deux épreuves, qui sont évoquées dans le paragraphe 5 : l'arbre. Ce dernier permet de représenter les différentes issues d'une expérience aléatoire, puis en le pondérant de faire apparaître les probabilités de chacune d'elles, comme le montrent les exemples suivants.

Situation	Arbre des possibles	Arbre pondéré avec les probabilités
Tirage à pile ou face		
Lancer d'un dé équilibré		



On exploite ce dernier exemple pour mettre en évidence l'emploi de l'arbre pondéré pour calculer la probabilité de l'événement "ne pas tirer une boule portant le numéro 1" :

- en additionnant les probabilités figurant sur les branches relatives aux résultats autres que 1 : on trouve 8/12, c'est-à-dire 2/3 ;
- en utilisant le fait que la somme des probabilités figurant sur l'ensemble des branches est égale à 1, et en prenant le complément à 1 de celle figurant sur la première branche.



Un traitement avec l'arbre permet de calculer la probabilité de tirer une boule portant un numéro pair, en additionnant les probabilités figurant sur les branches relatives à un numéro pair.

4. Langage et propriétés

À partir des exemples traités, quelques éléments de langage et propriétés sont institutionnalisés, en employant le langage des événements.

• Deux événements sont incompatibles s'ils ne peuvent se produire en même temps.
• L'événement contraire d'un événement est celui qui se réalise lorsque l'événement n'a pas lieu.
• La probabilité d'un événement est comprise entre 0 et 1. On peut l'exprimer sous diverses

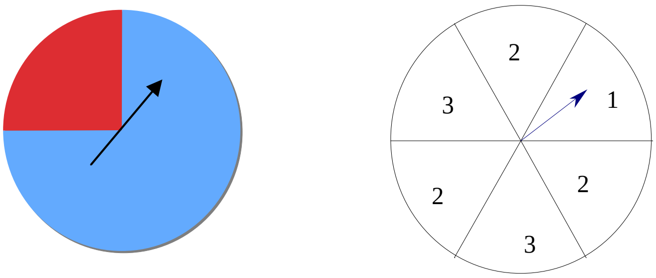
formes (décimale, fractionnaire, pourcentage).
• La probabilité d'un événement qui se produit nécessairement (événement certain) est égale à 1.
• Si deux événements sont incompatibles, la probabilité que l'un ou l'autre se réalise est égale à la somme de leurs probabilités. Plus généralement, on peut additionner les probabilités d'événements deux à deux incompatibles.
• Un événement et son contraire sont incompatibles et la réalisation de l'un ou de l'autre est certaine. Donc la somme de leurs probabilités est égale à 1.
En particulier, la probabilité d'un événement qui ne peut pas se produire (événement impossible) est égale à 0.

Pour faciliter les échanges, les événements (obtenir « Face », obtenir un nombre impair, …) peuvent être désignés par des lettres ou des symboles (ici, par exemple, la lettre F, la lettre I). La probabilité d'obtenir « Face » peut alors être notée *p*(obtenir « Face ») ou *p*(Face) ou *p*(F).

5. Expériences à deux épreuves

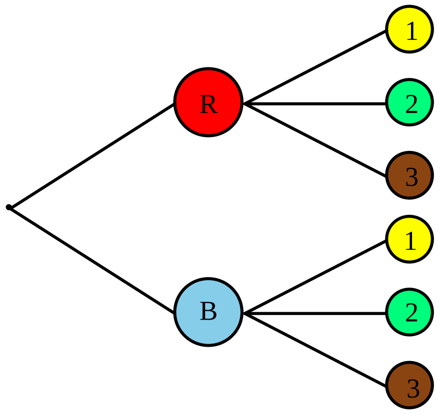
Aucune compétence sur les expériences à deux épreuves n'est exigible dans le cadre du socle commun. Les situations proposées aux élèves doivent rester élémentaires. On se bornera à des expériences conduisant à un maximum de 6 cas (2×2, 2×3 ou 3×2) et on n'abordera pas les cas de tirages successifs dans une urne (avec ou sans remise)[9].

On considère l'expérience suivante, qui se déroule en deux étapes : d'abord, on fait tourner la roue de loterie située ci-dessous à gauche (on obtient la couleur « Rouge » avec une probabilité de 0,25 et la couleur « Bleu » avec une probabilité de 0,75). Ensuite, on fait tourner la deuxième roue de loterie (on obtient le numéro 1 avec la probabilité 1/6, le numéro 2 avec la probabilité 1/2 et le numéro 3 avec la probabilité 1/3).

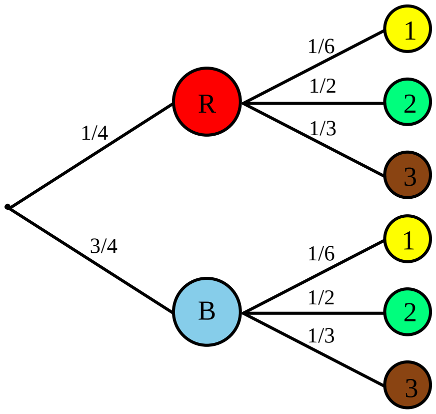


Un arbre non pondéré permet de déterminer tous les résultats possibles à l'issue de ces deux étapes. Les résultats possibles peuvent être notés ainsi : (R, 1), (R, 2), (R, 3), (B, 1), (B, 2), (B, 3). Chacun de ces résultats est représenté dans l'arbre ci-dessous par la succession de deux branches[10].

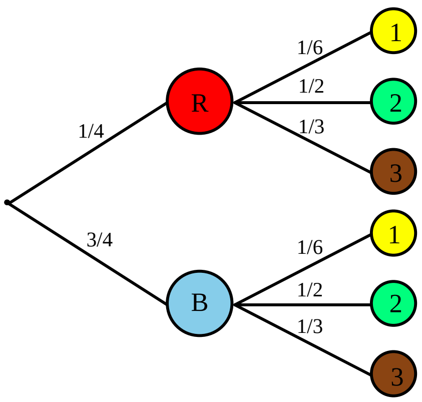
[9] Sont seulement envisageables le tirage d'une boule dans une urne, le tirage d'une boule dans une urne suivi du tirage d'une boule dans une autre urne, situations très proches des exemples proposés ci-dessous.
[10] Une telle succession de branches est souvent appelée "chemin".

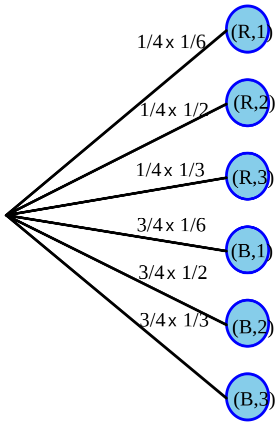


On cherche la probabilité d'obtenir chacun des six résultats possibles à l'issue des deux étapes : (R, 1), …, (B, 3).
On peut pondérer l'arbre avec les probabilités :

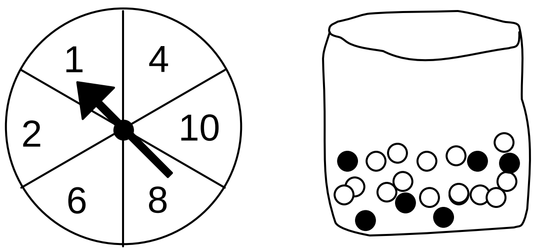


Comment évaluer la probabilité du résultat (R, 1) ?
L'approche fréquentiste de la probabilité permet la justification suivante :
Imaginons que l'on reproduise 120 000 (ou *N*) fois l'expérience. 1/4 de ces expériences environ suivront la branche vers R, et parmi celles-ci 1/6 environ iront vers 1.
Donc il y en a environ $\frac{1}{6}\left(\frac{1}{4}\times 120\,000\right)$ ou $\frac{1}{6}\left(\frac{1}{4}\times N\right)$, soit 5 000 (ou *N*/24) qui conduiront au résultat (R,1). La fréquence correspondante est 5 000/120 000 ou 1/24.
De manière plus générale, ceci conduit à admettre que la probabilité d'obtenir R à la première épreuve et 1 à la deuxième est égale au produit des probabilités 1/4 et 1/6, rencontrées sur le chemin représentant le résultat (R, 1).
Ce résultat peut être institutionnalisé, par exemple sous la forme suivante :
Dans un arbre, la probabilité du résultat auquel conduit un chemin est égal au produit des probabilités rencontrées le long de ce chemin.
Mais cette connaissance n'est pas un objectif du programme et on ne proposera que des exemples très simples dans lesquels un raisonnement permet facilement de trouver les résultats en leur donnant du sens.





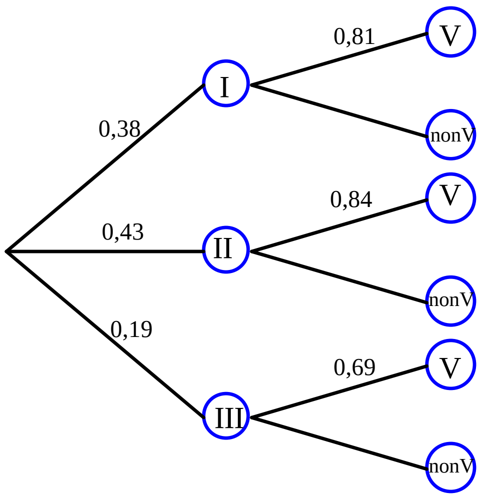
L'exercice suivant, inspiré de l'évaluation PISA 2003, peut être utilisé de diverses manières pour l'étude des expériences à deux épreuves :

> Un stand à la foire du printemps propose un jeu dans lequel il faut d'abord faire tourner une roulette. Ensuite, si la roulette s'arrête sur un nombre pair, le joueur peut tirer une bille dans un sac. La roulette et le sac de billes sont représentés ci-dessous : Des prix sont distribués aux joueurs qui tirent une bille noire. Suzy tente sa chance une fois.
>
> 
>
>
> Concernant le fait que Suzy gagne un prix, peut-on dire que :
> A C'est impossible.
> B C'est peu probable.
> C Il y a environ une chance sur deux.
> D C'est très probable.
> E C'est certain.

L'exemple suivant, cité dans [12], montre l'intérêt d'une représentation par un arbre dans une situation de la vie courante. Reposant essentiellement sur des calculs de pourcentages (pourcentages de pourcentages), cet exercice conforte l'explication proposée dans l'exemple introductif pour le calcul de (R, 1). On peut proposer, comme variante, de déterminer les effectifs de chaque catégorie d'électeurs, en donnant le nombre d'habitants de la ville.

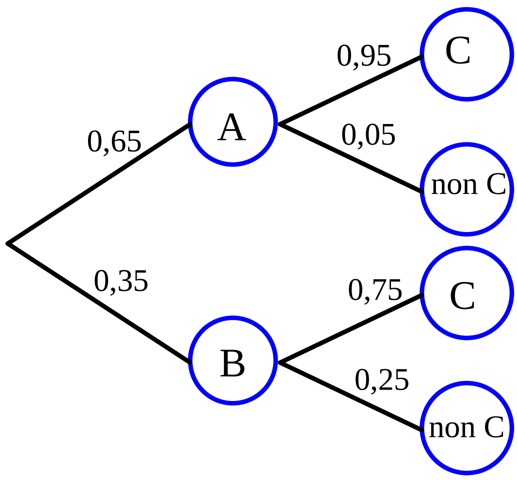
> Un scrutin a été organisé pour renouveler le conseil municipal d'une ville. Pour l'analyse des résultats, on distingue d'une part les électeurs, c'est-à-dire les personnes qui ont le droit de vote, d'autre part les votants, c'est-à-dire les personnes qui ont effectivement pris part au vote. De plus, pour cette analyse du scrutin, les électeurs sont répartis en trois groupes, en fonction de leur âge :
> - le groupe I, comprenant les électeurs de moins de 35 ans, représente 38% de l'ensemble des électeurs ;
> - le groupe II, comprenant les électeurs de 35 à 60 ans, représente 43% de l'ensemble des électeurs ;
> - le groupe III, comprenant les électeurs de plus de 60 ans, représente 19% de l'ensemble des électeurs.
>
> Enfin, les taux de participation ont pu être déterminés dans chacun des groupes : 81% dans le groupe I, 84% dans le groupe II, 69% dans le groupe III.
> On choisit au hasard un électeur. Quelle est la probabilité pour qu'il ait voté ?

L'énoncé permet de construire l'arbre pondéré ci-dessous :



Pour trouver la probabilité en question, il suffit d'additionner les probabilités des chemins qui aboutissent à V :
$0{,}38 \times 0{,}81 + 0{,}43 \times 0{,}84 + 0{,}19 \times 0{,}69$
c'est-à-dire environ 80%.

Un autre type d'exemples liés à la vie courante concernant la qualité de produits manufacturés permet de réinvestir les méthodes précédentes.
Un même produit manufacturé P est fabriqué dans deux usines A et B. Sur le marché concerné, l'usine A assure p% de la production et l'usine B ($1 - p$%). Une enquête a permis d'établir que q% des produits fabriqués par A satisfont aux normes, le pourcentage correspondant pour l'usine B étant de r%. On demande quelle est la probabilité qu'un produit P acheté au hasard soit conforme aux normes.
L'arbre suivant correspond aux valeurs suivantes de p, q et r : 65, 95, 75.



Cet arbre permet de répondre à la question.
Pour ce type d'exemple, il pourra être proposé, en variante, de déterminer les effectifs de chaque branche en donnant un nombre total assez grand d'objets manufacturés.

Exemple de situation se prêtant à expérimentation : (d'après le site www.statistix.fr):
Art et hasard : Des enfants réalisent des tableaux aléatoires avec des gommettes bleues, jaunes et vertes. Pour cela ils lancent plusieurs fois deux dés :
À chaque lancer

- si les deux faces obtenues sont impaires, ils collent une gommette bleue ;
- si les deux faces obtenues sont paires, ils collent une gommette jaune ;
- si les deux faces sont de parités différentes, ils collent une gommette verte.

Que penser des proportions de gommettes de chaque couleur ?

Après un temps d'expérimentation nécessaire à l'appropriation de la situation, on peut éventuellement simuler au tableur le tirage aléatoire :

	A	B	C	D	E
1	Tirage 1°	Tirage 2	Parité du dé 1	Parité du dé 2	Couleur gommette
2	=ALEA()	=ALEA()	=SI(A2<0,5;"I";"P")	=SI(B2<0,5;"I";"P")	=SI(C2=D2;SI(C2="I" ; "B" ;"J");"V")
3	=ALEA()	=ALEA()	=SI(A3<0,5;"I";"P")	=SI(B3<0,5;"I";"P")	=SI(C3=D3;SI(C3="I" ; "B" ;"J");"V")

On peut ensuite expliquer les fréquences observées en numérotant les deux dés pour représenter la situation aléatoire par un arbre dont les chemins sont les résultats du dé 1 puis du dé 2 : (I,I), (I,P), (P,I), (P,P).

On trouvera ci-dessous deux exercices proposés dans l'évaluation PISA, permettant d'évoquer avec les élèves des idées fausses concernant les jeux de hasard.

LOTERIE

Pour une loterie nationale, six boules sont tirées au hasard chaque semaine parmi quarante boules identiques numérotées de 1 à 40. Les gagnants du gros lot sont les joueurs qui ont choisi les six numéros tirés.

Le montant total du prix est partagé entre les gagnants.

Un journal publie les numéros gagnants de la semaine précédente ainsi qu'une liste des numéros qui ne sont plus sortis depuis longtemps.

Entourez soit « Vrai », soit « Faux » pour chacune des affirmations ci-dessous :

Affirmation	Vrai ou Faux
Les informations publiées par le journal ne sont d'aucune utilité pour prévoir les numéros de la semaine suivante, parce que toutes les combinaisons de six numéros ont autant de chances de sortir.	Vrai / Faux
Les numéros de la semaine précédente ont **davantage** de chances de sortir parce qu'ils sont « chauds ».	Vrai / Faux
Les numéros de la semaine précédente ont **moins** de chances de sortir parce qu'il est peu probable qu'un numéro sorte deux fois de suite.	Vrai / Faux
Les numéros qui ne sont plus sortis depuis longtemps ont **davantage** de chances de sortir.	Vrai / Faux

PILE OU FACE

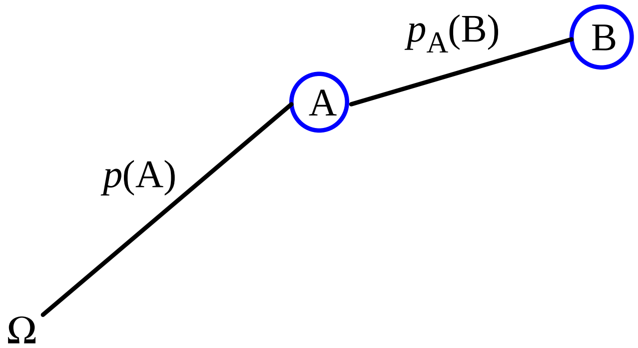
On a lancé 4 fois de suite une pièce de monnaie non truquée et chaque fois le résultat a été face.

Si on lance la même pièce une fois de plus, laquelle des affirmations suivantes sera correcte ?

A On a autant de chances d'obtenir pile que face.
B On a plus de chances d'obtenir pile.
C On a plus de chances d'obtenir face.
D On ne peut pas obtenir à nouveau face.

6. Continuité avec l'enseignement au lycée

L'arbre est un mode de représentation et de traitement que les élèves retrouveront s'ils poursuivent leur scolarité[11]. Il permet de traiter les questions faisant intervenir des probabilités conditionnelles. De manière générale, un chemin (succession de plusieurs branches) tel que ceux qui ont été utilisés précédemment sera interprété de la manière suivante dans le cadre de la théorie des probabilités :



Au bout du chemin, se trouve une feuille (qui n'est pas toujours indiquée), qui représente l'événement "A et B". Sa probabilité est le produit des probabilités rencontrées sur les branches le long du chemin : $p(\text{A et B}) = p(\text{A}) \times p_{\text{A}}(\text{B})$.

Puisque "A et B" et "B et A" désignent le même événement, on a également :

$$p(\text{A et B}) = p(\text{B}) \times p_{\text{B}}(\text{A}).$$

Le cas où les événements A et B sont (stochastiquement) indépendants est un cas particulier important. On a alors $p(\text{A et B}) = p(\text{A}) \times p(\text{B})$.

[11] Pour davantage de détails sur l'emploi des arbres, voir [12], [13] et [14] et plus généralement sur l'enseignement des probabilités au lycée, voir [15].

Éléments de bibliographie :

[1] Kahane J.-P. (dir.), 2002, *Rapport au ministre de l'Education nationale, L'enseignement des sciences mathématiques*, Commission de réflexion sur l'enseignement des mathématiques, CNDP & Éditions Odile Jacob.

[2] Hacking I., 2002, *L'émergence de la probabilité*, Editions du Seuil.

[3] Hacking I., Dufour M., 2004, *L'ouverture au probable, Eléments de logique inductive*, Armand Colin.

[4] HENRY M., 1994, L'enseignement du calcul des probabilités dans le second degré : perspectives historiques, épistémologiques et didactiques, *Repères IREM* N° 14.

[5] PICHARD J.-F., 1998, Approche épistémologique et diverses conceptions de la probabilité, *Repères IREM* N° 32.

[6] de Finetti B., 1937, La prévision : ses lois logiques, ses sources subjectives, Annales de l'I.H.P., Tome 7, n°1, 1-68. Téléchargeable sur le site : http://www.numdam.org/item?id=AIHP_1937__7_1_1_0.

[7] Picard P., 2007, *Hasard et probabilités. Histoire, théorie et application des probabilités*, Vuibert.

[8] Engel A., 1975, *L'enseignement des probabilités et de la statistique*, Cedic.

[9] GIRARD J.-C., HENRY M., PICHARD J.-F., 2001, Quelle place pour l'aléatoire au collège ? *Repères IREM* N° 42.

[10] HENRY M., 2003, Des lois continues, pourquoi et pour quoi faire ? *Repères IREM* N° 51.

[11] DUTARTE P., 2006, *L'induction statistique au lycée*, Didier.

[12] PARZYSZ B, 1990, Un outil sous-estimé : l'arbre probabiliste, Bulletin de l'APMEP n° 372, pp. 47-52.

[13] PARZYSZ B., 1993, *Des statistiques aux probabilités : exploitons les arbres*, Repères IREM n° 10.

[14] GRANGE J.-P., Arbres et tableaux en probabilités conditionnelles, in *Probabilités au lycée,* Commission Inter-IREM "Statistique et Probabilités", Brochure APMEP n° 143, 2003.

[15] HENRY M., 1999, L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie, *Repères IREM* N° 36.

Annexe 1 : Différentes interprétations de la probabilité

La théorie moderne des probabilités s'édifie axiomatiquement ; l'exposé complet en a été donné en 1933 par Kolmogorov. Par ailleurs, le rôle de la combinatoire dans la naissance du calcul des probabilités (travaux de Pascal par exemple) a beaucoup été étudié et vulgarisé. Cependant l'émergence du concept de probabilité ne peut se réduire à la résolution de problèmes issus de jeux de hasard, et l'axiomatisation de la théorie ignore volontairement l'interprétation de la probabilité et l'utilisation qui peut en être faite. Des travaux récents (Voir [2], [3] et [5]), notamment ceux de Ian Hacking, professeur au Collège de France[12], montrent la complexité de cette évolution, faisant apparaître essentiellement deux interprétations de la probabilité : en terme de croyance ou bien en terme de fréquence[13]. Dans la première, que Hacking nomme « probabilité épistémique », le mot « probable » est associé aux notions de croyance, confiance, crédibilité, indice ; dans la deuxième, nommée « probabilité de type fréquentiste », son usage est associé aux notions de fréquence, disposition, tendance, symétrie, propension. Ces deux interprétations de la probabilité sont parfois dénommées autrement : subjective / objective ; épistémique / aléatoire ; … [14] La description de la première peut être affinée en distinguant la probabilité personnelle, la probabilité interpersonnelle, et la probabilité logique. On peut illustrer ces différentes interprétations en prenant l'exemple d'une loterie de 1000 boules numérotées[15].

La probabilité personnelle

> En ce qui me concerne, je ne vois pas pourquoi sortirait une boule plutôt qu'une autre. Et donc, la probabilité de tirer l'une quelconque des boules est de 1/1000.

La probabilité interpersonnelle (*ou personnaliste)*

> Aucune personne sensée ne donnerait une probabilité plus forte à une boule qu'à une autre. Et donc, la probabilité de tirer l'une quelconque des boules est de 1/1000.

L'approche personnaliste est une variante de l'approche épistémique qui est opposée à une autre variante :

La probabilité logique

> La relation logique entre l'hypothèse h_j (hypothèse de tirage de la boule n°*j*) et les données expérimentales est la même que celle existant entre toute autre hypothèse h_j et ces mêmes faits.
> Les probabilités sont donc identiques. Et donc, la probabilité de tirer l'une quelconque des boules est de 1/1000.

La probabilité logique est sensée être une relation entre des informations ou indices et une hypothèse. Lorsqu'il ne dispose d'aucune information ou indice, le partisan de la probabilité logique utilise un principe appelé principe d'indifférence ou principe de raison insuffisante :

Le principe d'indifférence ou de raison insuffisante

> Il n'y a aucun indice permettant de privilégier l'une des 1000 hypothèses qui sont exclusives et conjointement exhaustives. Il faut donc attribuer à chacune d'elles la probabilité 1/1000.

On peut affiner de même l'interprétation fréquentiste, en distinguant deux interprétations opposées : celle de la "fréquence limite" et celle de propension, que l'on peut illustrer ainsi, à l'aide du même exemple :

La probabilité comme "fréquence limite"

> Lors de tirages répétés – avec remise – chaque boule est tirée aussi souvent qu'une autre. Et donc la probabilité de tirer l'une quelconque des boules est la fréquence relative du tirage de cette boule, à savoir 1/1000.

Cette interprétation repose sur la notion de fréquence relative "à long terme", idéalisée en

---

[12] Chaire de philosophie et histoire des concepts scientifiques.

[13] Voir [3], chapitres 11 et 12.

[14] Les différentes interprétations de la probabilité (et les dénominations variées les concernant) sont abordées dans [4] et [5].

[15] Les exemples qui suivent sont tirés de [3], page 160.

considérant une suite infinie d'essais dont l'indépendance est postulée. Elle insiste plus sur les résultats que l'on peut voir apparaître que sur les causes sous-jacentes à ces résultats, même si ce sont ces causes qui sont à l'origine de la stabilité des fréquences relatives. C'est précisément l'objet de l'interprétation en termes de tendance ou de propension, que l'on peut illustrer ainsi :

La probabilité comme propension

> L'urne, les boules, et la procédure de tirage sont agencées de telle sorte que la propension d'une boule à être tirée est la même que celle d'une autre. Et donc, la probabilité de tirer l'une quelconque des boules est de 1/1000.

Cette dernière interprétation est pertinente pour la modélisation de phénomènes naturels tels que la radioactivité.

La conception numérique de la probabilité date des années 1650. Pour en faire comprendre l'intérêt, on fait alors appel aux loteries et jeux de hasard, alors très en vogue, et ce que l'on entendait par « probabilité » ne posait guère problème. Il n'y a pas vraiment lieu de distinguer les deux grandes interprétations de la probabilité, dans un exemple typique comme celui qui précède.

Au XVIII^{e} siècle, on n'hésite pas à faire usage du calcul des probabilités dans des contextes géométriques, le problème type étant l'évaluation « des chances que l'on a de toucher une cible visée avec un projectile ». Les conditions du tirage au sort y sont moins bien définies que dans les jeux de hasard, mais ces contextes offrent des ensembles d'issues qui ne sont pas dénombrables : une probabilité est définie alors comme un rapport de longueurs, d'aires ou de volumes. Le problème le plus connu est sans doute celui de l'aiguille de Buffon (1733) :

> On jette « au hasard » sur un plan – le sol par exemple – portant des droites parallèles formant des bandes de largeur d, une aiguille de longueur l ($l < d$). On demande la probabilité de rencontre entre l'aiguille et l'une des droites parallèles.

Le résultat, $\frac{2l}{\pi d}$, peut être obtenu en calculant un rapport d'aires, l'une d'elles nécessitant le recours au calcul intégral[16]. Un autre exemple, accessible en classe de 3^{e}, est le jeu du « Franc Carreau » :

> On jette « au hasard » une pièce de diamètre d sur un sol pavé avec des carrés dont les côtés sont de longueur a, $a > d$. On demande la probabilité de voir la pièce tomber entièrement à l'intérieur d'un seul carré.

Sa solution conduit à calculer le rapport des aires de deux carrés concentriques, de côtés respectifs $a - d$ et a : il est égal à $\frac{(a-d)^2}{a^2}$.

[16] On peut également le démontrer en s'intéressant à l'espérance mathématique E(l) du nombre de points de rencontre, en ignorant la restriction $l < d$. En utilisant la linéarité de l'espérance, il suffit de déterminer le nombre k tel que E(l) = kl, ce que l'on peut faire en considérant une aiguille de forme circulaire de diamètre d, pour laquelle on sait que l'espérance vaut 2. Cette méthode montre la puissance d'une autre approche des probabilités (Huygens et Van Schooten - 1657) reposant sur la notion d'espérance mathématique. Elle a été exploitée par Laplace, qui a généralisé le résultat de Buffon, en remplaçant les droites parallèles par un réseau à maille rectangulaire. En comptant le nombre de points d'intersection d'une courbe avec un tel réseau, on peut ainsi estimer cette espérance, et en déduire la longueur de la courbe.

Annexe 2 : Éléments d'histoire de la notion de probabilité

Comme on l'a vu dans l'annexe 1, il n'y a guère lieu de distinguer les deux grandes interprétations de la probabilité (la probabilité épistémique et la probabilité de type fréquentiste) dans l'exemple typique du tirage de boules dans une urne. Mais ces deux grandes interprétations refont surface quand on considère des exemples plus quotidiens (ainsi que dans la modélisation de situations plus complexes). Elles interviennent également dans l'enseignement des probabilités. C'est la raison pour laquelle ces deux interprétations sont présentées succinctement ci-dessous : la probabilité épistémique fait l'objet du paragraphe 1., la probabilité de type fréquentiste est traitée au 2.

1. Probabilité épistémique

On considère un événement E, et deux personnes qui ne savent pas si E va ou non se réaliser. La première personne I parie x € sur E et la deuxième personne J parie y € contre E : si I gagne le pari, elle empoche les y € de J ; si J gagne, elle empoche les x € de I. La somme $(x + y)$€ est appelée enjeu du pari, $x/(x + y)$ est le taux de pari de I sur l'événement E, et $y/(x + y)$ est le taux de pari de J contre l'événement E. Le taux de pari est égal à la mise divisée par l'enjeu.

Les joueurs ne parlent pas de taux de pari, mais plutôt de chance. Si les chances contre un événement E sont de b contre a, le taux de pari sur E est de $a/(a + b)$. Parler de taux de pari ou de chance sont deux façons d'exprimer la même chose. Mais la notion de taux de pari ressemble davantage à la probabilité. Cependant, alors qu'un joueur qui parie escompte un gain[17], une probabilité (personnelle ou épistémique) est un taux de pari *équitable* p : la personne concernée estime qu'il est indifférent de parier au taux p sur E ou de parier au taux $(1 – p)$ contre E. S'il s'agit de définir des taux de pari relatifs à plusieurs événements (p_1 pour E_1, p_2 pour E_2, …), le mathématicien italien Bruno de Finetti[18] explique dans les termes qui suivent comment on peut justifier les règles de calcul sur les probabilités :

> Lorsqu'un individu a évalué les probabilités de certains événements, deux cas peuvent se présenter : ou bien il est possible de parier avec lui en s'assurant de gagner à coup sûr, ou bien cette possibilité n'existe pas. Dans le premier cas, on doit dire évidemment que l'évaluation de la probabilité donnée par cet individu contient une incohérence. C'est précisément cette condition de cohérence qui constitue le seul principe d'où l'on puisse déduire tout le calcul des probabilités : ce calcul apparaît alors comme l'ensemble des règles auxquelles l'évaluation subjective des probabilités de divers événements par un même individu doive être assujettie si l'on ne veut pas qu'il y ait entre elles une contradiction fondamentale.

Il démontre ensuite que si E_1, E_2, …, E_n constituent ce que l'on nomme aujourd'hui un système complet d'événements, p_1, p_2, …, p_n étant leurs probabilités respectives, la cohérence oblige à imposer la condition $p_1 + p_2 + \ldots + p_n = 1$.

Le lecteur trouvera dans [4] la démonstration de toutes les règles du calcul des probabilités à partir de cette notion de cohérence[19], y compris la formule de Bayes, qui joue un rôle très important dans les applications de la notion de probabilité épistémique, comme le montre l'exemple ci-dessous également tiré de [3].

---

[17] Un joueur I qui pense que "Non E" a trois fois plus de chance de se produire que E, misera 1€ sur E en espérant gagner plus que 3 € : $(3 + x)$€. Son taux de pari sur E est donc $1/(1 + 3 + x)$. Son adversaire J qui mise 3 € contre E, veut gagner plus que 1 € : $(1 + y)$€ , il offre 3 € pour gagner $(1 + y)$€ au cas où E ne se produit pas. Son taux de pari contre E est donc de $3/(1 + 3 + y)$. La somme des deux taux de pari est strictement inférieure à 1.

[18] Voir [6] en bibliographie. La première théorie systématique de la probabilité épistémique a été présentée en 1926 par Franck Ramsey. Les travaux de de Finetti datent de 1930 (alors que ceux de Kolmogorov datent de 1933).

[19] Voir [3], chapitre 14, intitulé "Cohérence".

Vous êtes médecin. Vous estimez tout à fait probable que l'un de vos patients ait une angine, mais vous n'en êtes pas sûr. Vous faites quelques prélèvements buccaux que vous envoyez au laboratoire pour analyse. Comme quasiment tous les tests, celui-ci n'est pas parfait.
Si le patient a une angine, dans 70% des cas le laboratoire dit « oui, il y a une angine ». Mais dans 30% des cas il dit qu'il n'y a pas d'angine.
Si le patient n'a pas d'angine, dans 90% des cas le laboratoire dit que le patient n'a pas d'angine. Mais dans 10% des cas, il prétend qu'il en a une.
Vous envoyez successivement cinq prélèvements (issus du même patient) pour analyse. Et voici les réponses : oui, non, oui, non, oui.
Qu'en concluez-vous :

(a) Ces résultats ne valent rien.
(b) Il est probable que le patient n'ait pas d'angine.
(c) Il est juste un peu plus probable qu'il ait une angine.
(d) Il est très nettement plus probable qu'il ait une angine.

On trouvera ci-dessous la justification du fait que la bonne réponse est (d) : en supposant que la probabilité que le patient ait une angine soit pour le médecin de 90% (probabilité antérieure ou *a priori*), la probabilité que le patient ait une angine sachant que les résultats aux cinq tests (supposés indépendants) sont “oui – non – oui – non – oui” (probabilité postérieure ou *a posteriori*) est égale à 343/344, soit environ 0,997. Si l'on remplace le médecin par une personne ignorante en la matière, cette dernière pourra fixer la probabilité a priori à 50%. Les mêmes calculs conduisent à une probabilité a posteriori de 343/352 soit environ 0,974.
Justification :

Désignons par A l'événement “Le patient a une angine”, par O le fait que le test soit positif, et par N le fait qu'il soit négatif.
On connaît les probabilités conditionnelles suivantes :
$p_A(\text{O}) = 0{,}7$; $p_A(\text{N}) = 0{,}3$; $p_{\overline{A}}(\text{O}) = 0{,}1$; $p_{\overline{A}}(\text{N}) = 0{,}9$.
On cherche la probabilité que le patient ait une angine (A) sachant que les résultats aux cinq tests successifs sont respectivement : O, N, O, N, O, que nous noterons $p_{\text{O et N et O et N et O}}(\text{A})$. On suppose que les résultats aux tests sont indépendants.
On a donc :
$p_A(\text{O et N et O et N et O}) = 0{,}7 \times 0{,}3 \times 0{,}7 \times 0{,}3 \times 0{,}7$
et $p_{\overline{A}}(\text{O et N et O et N et O}) = 0{,}1 \times 0{,}9 \times 0{,}1 \times 0{,}9 \times 0{,}1$.
En appliquant la formule de Bayes, on obtient :

$$p_{\text{O et N et O et N et O}}(\text{A}) = \frac{p(\text{A}) \times p_A(\text{O et N et O et N et O})}{p(\text{A}) \times p_A(\text{O et N et O et N et O}) + p(\overline{\text{A}}) \times p_{\overline{A}}(\text{O et N et O et N et O})}$$

Le médecin prend 0,9 comme valeur pour $p(\text{A})$.
On trouve alors que $p_{\text{O et N et O et N et O}}(\text{A})$ vaut 343/344, soit 0,997 environ.

La personne ignorante préfère prendre 0,5 comme valeur pour $p(\text{A})$. Les calculs conduisent alors au résultat 343/352, c'est-à-dire 0,974 environ.

Rappels :
• La formule suivante :
$p(\text{B}) = p_A(\text{B}) \times p(\text{A}) + p_{\overline{A}}(\text{B}) \times p(\overline{\text{A}})$
est appelée formule des probabilités totales.
• La formule de Bayes permet de calculer $p_B(\text{A})$ connaissant $p_A(\text{B})$ et $p_{\overline{A}}(\text{B})$.
En effet : $p_B(\text{A}) = \dfrac{p(\text{B et A})}{p(\text{B})}$. Or $p(\text{B et A}) = p_A(\text{B}) \times p(\text{A})$ et d'après la formule des probabilités totales : $p(\text{B}) = p_A(\text{B}) \times p(\text{A}) + p_{\overline{A}}(\text{B}) \times p(\overline{\text{A}})$. Donc :

$$p_{\mathrm{B}}(\mathrm{A}) = \frac{p_{\mathrm{A}}(\mathrm{B}) \times p(\mathrm{A})}{p_{\mathrm{A}}(\mathrm{B}) \times p(\mathrm{A}) + p_{\overline{\mathrm{A}}}(\mathrm{B}) \times p(\overline{\mathrm{A}})}.$$

Cette formule se généralise en remplaçant le système complet d'événements $\{\mathrm{A}, \overline{\mathrm{A}}\}$ par un système complet (ou partition) formés de *n* événements, souvent notés H_1, H_2, …, H_n.

Dans cet exemple, la probabilité de type épistémique est clairement mise en jeu.

2. Probabilité de type fréquentiste

Cette interprétation de la probabilité est beaucoup plus familière, car tout étudiant en mathématiques en entend parler lorsqu'on traite des lois des grands nombres.

2.1 Le théorème de Bernoulli

Du point de vue historique, le résultat fondamental permettant de comprendre la façon dont se stabilisent les fréquences des résultats au fur et à mesure que le nombre des essais augmente est le théorème de Bernoulli[20]. Ce théorème, qui confirme l'intuition sur la stabilité des fréquences, est plus difficile à démontrer que la formule de Bayes. La démonstration "moderne" dans la théorie axiomatique de Kolmogorov repose sur une majoration assez grossière donnée par la formule de Bienaymé - Tchebychev, majoration peu utile en pratique. La démonstration donnée par Bernoulli est beaucoup plus compliquée, mais présente l'intérêt de présenter très clairement sa problématique : il remarque que, dans les jeux de dés ou de tirages dans une urne, la détermination des probabilités a priori ne pose pas de problème : il suffit de prendre le ratio entre le nombre de tirages "fertiles" et le nombre total de tirages ou le ratio entre le nombre de tirages "fertiles" et le nombre de tirages "stériles". Mais, il constate que cette méthode est inutilisable dans des problèmes concernant les maladies, la météorologie, où les causes sont cachées, et où l'énumération des cas équiprobables est impossible. Au lieu de cela, il propose de déterminer la probabilité d'un cas favorable a posteriori :

> « On peut supposer qu'une chose particulière se produira ou non autant de fois qu'elle s'est produite ou non dans le passé, dans des circonstances semblables ».

Il cherche donc à déterminer empiriquement la proportion de cas favorables dans le cas où elle est inconnue. L'originalité de la tentative de Bernoulli consiste à donner un traitement formel de la vague notion qu'il décrit ainsi :

> « Même le plus stupide des hommes, par quelque instinct de la nature, par lui-même et sans aucune instruction (et c'est une chose remarquable), est convaincu que plus on fait d'observations, moins on risque de s'écarter de notre but ».

Bernoulli veut démontrer ce principe, et montrer que la "certitude morale" à propos de la proportion inconnue peut être approchée d'aussi près que l'on veut. Il considère une urne contenant 3000 galets blancs et 2000 galets noirs. On tire un galet, avec remise. On regarde combien de fois on tire un galet blanc, combien de fois on tire un galet noir. Se pose alors la question : Peut-on tirer un nombre suffisant de fois de manière à ce qu'il devienne 10 fois, 100 fois, 1000 fois plus probable que les nombres de galets blancs et noirs tirés soient dans le ratio 3:2 plutôt que dans tout autre ratio ? Bernoulli précise :

> « Pour éviter les malentendus, on doit noter que le ratio que nous essayons de déterminer expérimentalement ne doit pas être considéré comme précis et indivisible (sinon, c'est le contraire

[20] Ce théorème est démontré par J. Bernoulli dans *Ars Conjectandi* (1713), alors que la formule de Bayes figure dans un texte posthume de Thomas Bayes intitulé *Essai en vue de résoudre un problème de la doctrine des chances*.

qui se produirait, et il deviendrait moins probable que le vrai ratio soit trouvé en augmentant les observations). Ce que l'on veut, en revanche, c'est un ratio pris avec quelque latitude, c'est-à-dire situé entre deux limites qui peuvent être aussi proches l'une de l'autre que l'on veut. Par exemple, on prend deux ratios 301:200 et 299:200 ou 3001:2000 et 2999:2000 … l'un qui est immédiatement supérieur et l'autre immédiatement inférieur au ratio 3:2. On prouvera que l'on peut rendre plus probable que le ratio trouvé après des expériences répétées tombe entre ces limites plutôt qu'il tombe à l'extérieur. »

Avec les notations modernes usuelles, où X désigne le nombre de cas "fertiles" et n le nombre d'expériences répétées, il veut démontrer :

$$P\left(\left|\frac{X}{n}-p\right|\leq\varepsilon\right)>cP\left(\left|\frac{X}{n}-p\right|>\varepsilon\right) \text{ avec } c=10{,}100 \text{ ou } 1000, \text{ ou encore : } P\left(\left|\frac{X}{n}-p\right|\leq\varepsilon\right)>\frac{c}{c+1},$$

avec $p=\dfrac{r}{r+s}$ où r désigne le nombre de galets blancs et s celui des noirs. Après avoir fait la démonstration[21], il traite l'exemple où $r = 30$, $s = 20$, $p = 3/5$ et $\varepsilon = 1/50$. Il trouve $n = 25\,550$ pour $c = 1000$, ce qui se traduit ainsi en termes modernes : $P\left(\left|\dfrac{X}{25550}-\dfrac{3}{5}\right|\leq\dfrac{1}{50}\right)>\dfrac{1000}{1001}$.

25 550 est à l'époque un nombre astronomique, inutilisable dans la pratique (Il est pourtant bien meilleur que celui obtenu en employant l'inégalité de Bienaymé Tchebychev, qui est égal à 600 600) : ceci conduit Bernoulli à ne pas publier ses travaux. Remarquons que Bernoulli ne répond pas à la question qu'il s'est posée au départ, car la proportion est ici connue au départ. Il détermine le nombre de tirages suffisant pour que, avec une probabilité très forte (supérieure à 0,999, alors qu'aujourd'hui on se contente selon les secteurs d'activité des niveaux 0,95 ou 0,99), la proportion de cas favorables ne s'écarte pas de 3/5 de plus de 1/50. En d'autres termes, il détermine n pour un intervalle de probabilité[22] au niveau 1000/1001 d'amplitude donnée, alors qu'il cherchait n pour en faire un intervalle de confiance[23] au même niveau.

Bayes s'est posé et a résolu un problème voisin de celui de Bernoulli, qu'il formule ainsi[24] :

« Etant donné le nombre de fois qu'un événement inconnu s'est réalisé ou non, on cherche la chance que la probabilité de sa réalisation lors d'une seule épreuve soit comprise entre deux degrés quelconques de probabilité que l'on puisse assigner. »

problème que l'on peut traduire avec un langage plus moderne de la manière suivante :

Un événement se produit à chaque tirage avec la probabilité θ. Soit X le nombre de fois qu'il se produit au cours de n essais. On demande $P(a < \theta < b \mid X)$, probabilité que θ soit comprise entre a et b, connaissant le nombre de fois où l'événement s'est produit au cours des n essais.

Il a illustré sa résolution avec un dispositif original et intéressant, connu sous le nom de "billard de Bayes". Mais ses travaux n'ont connu aucune diffusion, en raison des difficultés de calcul des intégrales mises en jeu dans la solution.

2.2 Théorèmes de convergence et fluctuation d'échantillonnage

21 On trouvera une version modernisée de cette démonstration dans [7], pages 115 à 117 et la version originale dans Barbin E. & Lamarche J.-P., 2004 (dir.), *Histoires de probabilités et de statistiques*, Ellipses, pages 121 à 140, article écrit par Michel Henry.

22 Voir le paragraphe 2.2 de cette annexe.

23 Voir le paragraphe 2.3 de cette annexe.

24 Voir le chapitre 1 de Droesbeke J -J., Fine J., Saporta G, 2002, *Méthodes bayésiennes en statistique*, Editions Technip. Les auteurs, remarquant que Bayes emploie simultanément les mots « probabilité » et « chance », expliquent que si la « probabilité » peut correspondre au même concept que celui envisagé par Bernoulli, la « chance » représente plutôt une « raison de croire ».

Les travaux de De Moivre sur le développement du binôme, de De Moivre et de Laplace sur le théorème qui porte leurs noms (et qui est un cas particulier du théorème “central limite” établi par Laplace) et ceux de Gauss sur la loi de Laplace - Gauss vont permettre d’obtenir les résultats essentiels et de prolonger le travail de Bernoulli (et de Bayes).

On peut modéliser n tirages au hasard et avec remise dans une urne contenant des boules noires et des boules blanches, la proportion de boules noires dans l’urne étant égale à p, en introduisant n variables aléatoires indépendantes X_i, prenant la valeur 1 si la $i^{\text{ème}}$ boule tirée est noire, et 0 si elle est blanche. La somme S_n des X_i, i variant de 1 à n, donne le nombre de boules noires obtenu à l’issue des n tirages. S_n suit la loi binomiale de paramètres n et p, son espérance mathématique est égale à np et son écart type est égal à $\sqrt{np(1-p)}$. On s’intéresse à la fréquence F_n du nombre de boules noires à l’issue des n tirages, égale à $\frac{S_n}{n}$.
Le théorème de Moivre - Laplace dit que la variable aléatoire centrée réduite R_n associée à S_n, égale à $\frac{S_n - np}{\sqrt{np(1-p)}}$, converge en loi vers la loi normale centrée réduite, c’est-à-dire que lorsque n tend vers l’infini, la probabilité pour que R_n prenne des valeurs comprises entre a et b tend vers $\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx$. La variable centrée réduite associée à F_n est $R_n = \frac{F_n - p}{\sqrt{\frac{p(1-p)}{n}}}$.
Plus précisément, sous les hypothèses suivantes : $n > 30$, $np > 5$ et $n(1 - p) > 5$, on approxime avec une très bonne précision la probabilité pour que R_n soit dans l’intervalle $[a, b]$ par sa limite donnée par le théorème de Moivre - Laplace.
À l’aide d’une table de la loi normale ou d’un tableur, on peut trouver une valeur approchée du nombre réel u tel que $\frac{1}{\sqrt{2\pi}}\int_{-u}^{u} e^{-\frac{x^2}{2}}dx$, intégrale notée $\phi(u)$, prenne une valeur donnée.
Il est utile de retenir que $\phi(1{,}96) \approx 0{,}95$; $\phi(1{,}65) \approx 0{,}90$; $\phi(3) \approx 0{,}99$. Par exemple, sous les conditions rappelées plus haut : $P(-1{,}96 \leq R_n \leq 1{,}96) \approx 0{,}95$. Or $-1{,}96 \leq R_n \leq 1{,}96$ équivaut à :

$$p - 1{,}96\sqrt{\frac{p(1-p)}{n}} \leq F_n \leq p + 1{,}96\sqrt{\frac{p(1-p)}{n}}.$$

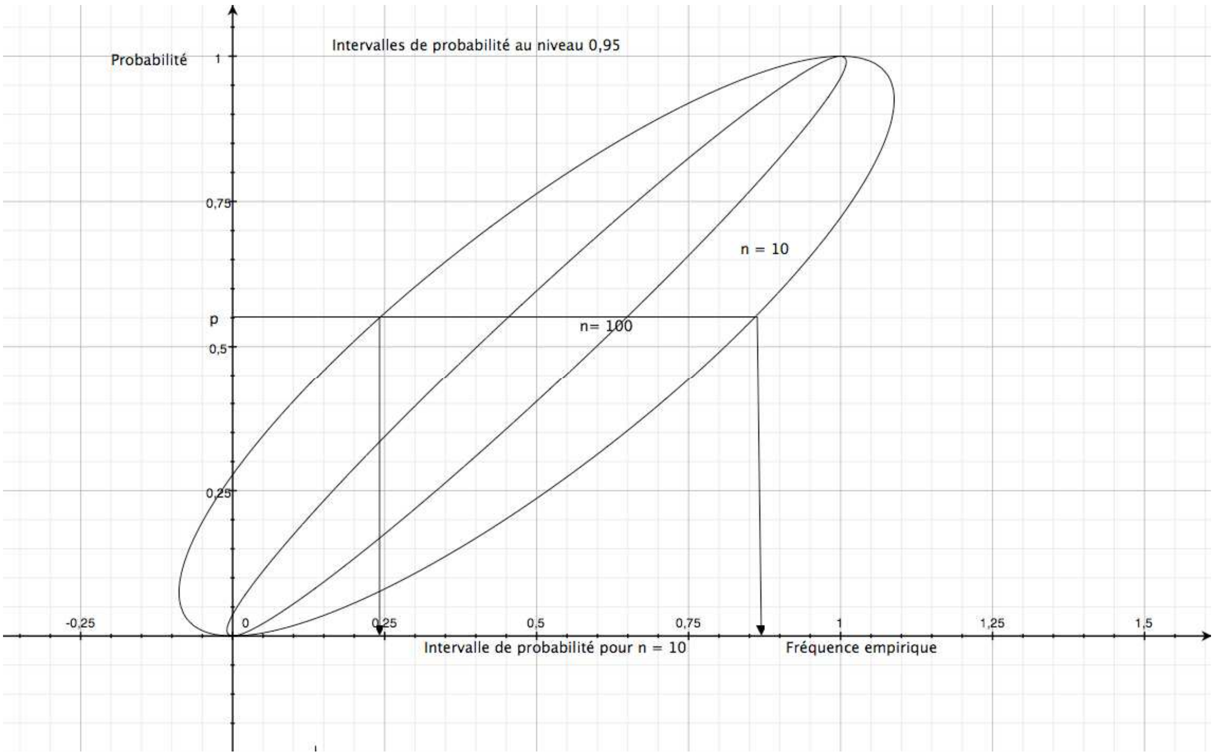
Donc avec une probabilité voisine de 95%, l’inégalité précédente est vraie. L’intervalle ayant pour extrémités $p - 1{,}96\sqrt{\frac{p(1-p)}{n}}$ et $p + 1{,}96\sqrt{\frac{p(1-p)}{n}}$ est appelé intervalle de probabilité (ou de pari) de niveau 95%, ou encore intervalle de fluctuation de niveau 95%. Son interprétation est la suivante : dans 95% des séries de n tirages que l’on peut faire, la fréquence empirique f_n obtenue expérimentalement (modélisée par F_n) appartient à cet intervalle.

Reprenons l’exemple traité par Bernoulli, dans lequel $p = 3/5$ et $\sqrt{\frac{p(1-p)}{n}} = \frac{\sqrt{6}}{5\sqrt{n}}$; il cherchait n tel que $P\left(\frac{3}{5} - 0{,}02 \leq F_n \leq \frac{3}{5} + 0{,}02\right) \approx 0{,}999$. En utilisant le théorème de Moivre – Laplace, on est conduit à chercher u tel que $\phi(u) = 0{,}999$. Une table de la loi normale donne $u \approx 3{,}29$: ainsi $P(-3{,}29 \leq R_n \leq 3{,}29) \approx 0{,}999$. Or $\frac{3}{5} - 0{,}02 \leq F_n \leq \frac{3}{5} + 0{,}02$

équivaut à $-0{,}02\frac{5\sqrt{n}}{\sqrt{6}} \le R_n \le 0{,}02\frac{5\sqrt{n}}{\sqrt{6}}$. Il suffit donc de choisir n tel que $0{,}02\frac{5\sqrt{n}}{\sqrt{6}} \approx 3{,}29$, c'est-à-dire $n \approx 6495$ (à comparer avec 25 550 obtenu par Bernoulli et à 600 600 obtenu avec l'inégalité de Bienaymé – Tchebychev).

2.3 Théorèmes de convergence et estimation d'une probabilité

Tout ce qui précède suppose que la probabilité p de succès lors d'un tirage est connue.
On peut donner une interprétation graphique de l'intervalle de probabilité (ou de pari), ou de fluctuation au niveau 95%. L'inégalité $-1{,}96\sqrt{\frac{p(1-p)}{n}} \le f_n - p \le +1{,}96\sqrt{\frac{p(1-p)}{n}}$ est équivalente à $(f_n - p)^2 \le \frac{1{,}96^2}{n}p(1-p)$, qui est l'équation de l'intérieur d'une ellipse[25] dans un repère où l'on porte f_n en abscisse et p en ordonnée : on lit l'intervalle de probabilité (ou de fluctuation) sur l'axe des abscisses, comme l'indique la représentation graphique ci-dessous. Elle permet par ailleurs de constater la réduction de l'intervalle lorsque n augmente.

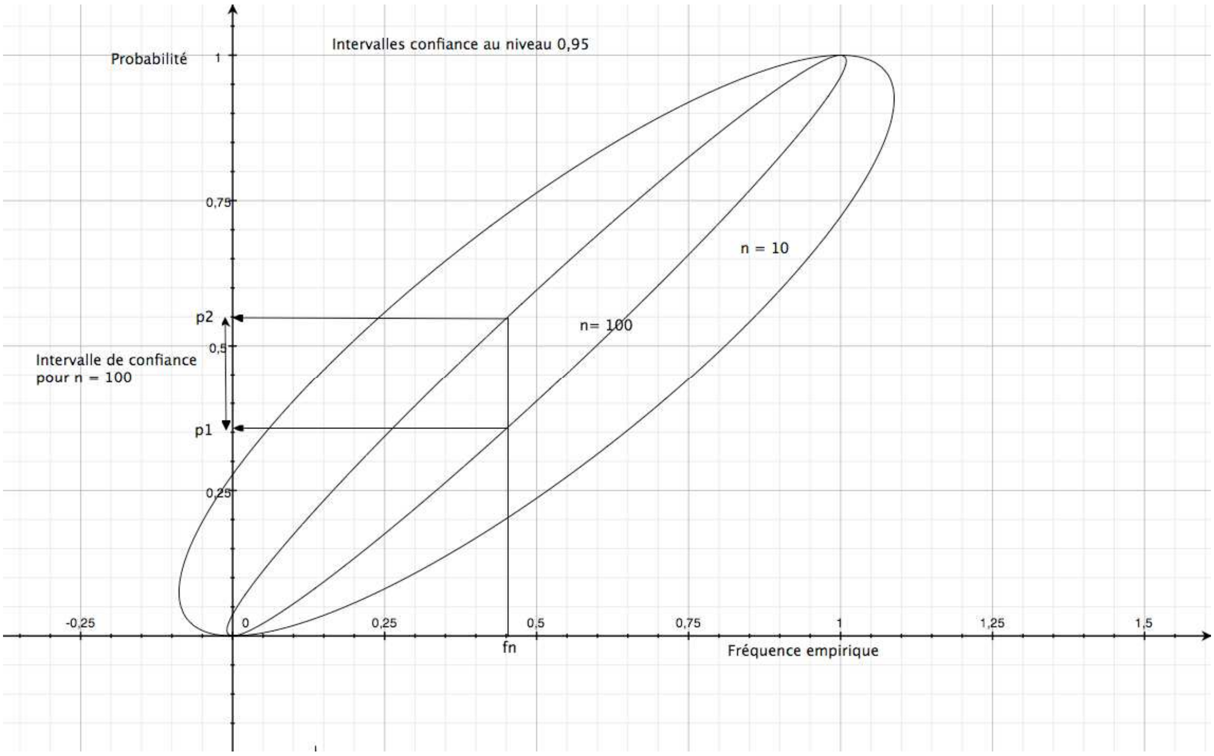


Le problème précédent fait partie de la théorie de l'échantillonnage, qui fournit, pour chaque valeur de p fixée, un intervalle de probabilité où la variable aléatoire prend ses valeurs avec une certaine probabilité (ici, 95%).

Ce problème est en quelque sorte l'inverse de celui dans lequel on cherche à estimer une probabilité à partir de la fréquence empirique obtenue lors d'une série de n tirages.
Ce problème fait partie de la théorie de Neyman[26], et sa solution est connue sous le nom d'intervalle de confiance à un niveau donné. En gardant le même niveau de 95%, le graphique

25 Seuls les points de l'intérieur de l'ellipse dont les deux coordonnées sont comprises entre 0 et 1 sont à prendre en considération.
26 Jerzy Neyman (1894 – 1981) a développé dans les années 1930 la théorie moderne des tests d'hypothèse avec Egon Pearson, fils de Karl Pearson, puis a développé ensuite celle des intervalles de confiance.

ci-dessous laisse deviner comment on le définit.



Dans le cas simple traité ici, il est possible de déterminer par le calcul les bornes p_1 et p_2 de l'intervalle de confiance au niveau 0,95. Il suffit pour cela de résoudre l'inéquation du second degré $(f_n - p)^2 \leq \frac{1{,}96^2}{n} p(1-p)$ d'inconnue p. L'intervalle $[p_1, p_2]$ obtenu est aléatoire, car ses bornes dépendent de f_n, qui lui-même l'est. Il est incorrect de dire que la proportion p de boules noires a une probabilité de 95% de se trouver dans l'intervalle $[p_1, p_2]$. En effet, une fois cet intervalle déterminé, ou bien il recouvre p, ou bien il ne la recouvre pas.

La probabilité est dans la procédure, pas dans le résultat. Autrement dit, l'estimation de p à l'aide de l'intervalle de confiance est faite avec une méthode dont la probabilité d'être juste est[27] de 95% : sur un grand nombre d'intervalles de confiance, obtenus à partir d'un grand nombre d'échantillons, 95% recouvrent effectivement la valeur p.

Il reste à déterminer les bornes p_1 et p_2 de l'intervalle de confiance. On sait que la probabilité pour que l'inégalité $p - 1{,}96\sqrt{\frac{p(1-p)}{n}} \leq f_n \leq p + 1{,}96\sqrt{\frac{p(1-p)}{n}}$ soit vraie est de 0,95 environ.

Cette inégalité est équivalente à $f_n - 1{,}96\sqrt{\frac{p(1-p)}{n}} \leq p \leq f_n + 1{,}96\sqrt{\frac{p(1-p)}{n}}$. Mais les nombres encadrant p dépendent de p ! On peut évidemment penser à remplacer dans leurs expressions p par f_n, c'est-à-dire remplacer la quantité inconnue p par la valeur observée f_n, comme l'a fait Neyman en présentant sa méthode, suscitant ainsi la méfiance de son collègue Bowley[28]. En fait, on peut légitimer cette pratique en montrant que les solutions de l'inéquation du second degré $(f_n - p)^2 \leq \frac{1{,}96^2}{n} p(1-p)$ sont très voisines de

[27] Cette probabilité est en fait de 95% environ, si l'on tient compte de l'approximation par la loi normale.
[28] Voir [11], pages 87 à 89.

$f_n - 1{,}96\sqrt{\dfrac{f_n(1-f_n)}{n}}$ et de $f_n + 1{,}96\sqrt{\dfrac{f_n(1-f_n)}{n}}$, à condition que n soit suffisamment grand (de l'ordre d'une centaine) et si p n'est ni trop grande, ni trop petite (np et $n(1-p)$ sont de l'ordre de 10 au plus).

Le programme de seconde propose une "fourchette de sondage au niveau de confiance 0,95", $\left[f_n - \dfrac{1}{\sqrt{n}}, f_n + \dfrac{1}{\sqrt{n}}\right]$, dont l'expression est plus simple, et qui contient l'intervalle de confiance au niveau 0,95 : en effet, $f_n(1-f_n)$ est inférieur à 1/4, et donc $1{,}96\sqrt{f_n(1-f_n)}$ est inférieur à 1. La probabilité qu'une telle fourchette recouvre la valeur p est donc supérieure à 0,95. Les conditions d'utilisation de cette fourchette ($n > 30$, et $0{,}3 < f_n < 0{,}7$) permettent d'assurer la validité de l'approximation par la loi normale et de s'assurer que la "fourchette" n'est pas trop différente de l'intervalle de confiance.