# LES PARADOXES DU PARADOXE DES ANNIVERSAIRES

**Stéphane VINATIER**[1]
IREM de Limoges

**Résumé.** À partir de l'exercice classique du paradoxe des anniversaires, nous discutons l'importance de la prise en compte de la démarche de modélisation dans l'enseignement des probabilités, du collège à l'université, pour donner leur vrai sens aux applications de la théorie des probabilités à des situations réelles et aux résultats qu'on obtient ainsi.

**Mots-clés.** Enseignement des mathématiques, probabilités, modélisation mathématique.

## *Introduction*

Le *paradoxe des anniversaires* est un exercice classique de probabilités élémentaires, qui demande de déterminer le nombre minimal de personnes dans un groupe pour que la probabilité qu'au moins deux d'entre elles aient la même date d'anniversaire soit supérieure à $\frac{1}{2}$. La résolution standard de l'exercice passe par l'utilisation d'une minoration de la fonction exponentielle pour une solution aprochée, ou par un peu de programmation pour une solution exacte. Elle repose sur des choix de modélisation le plus souvent implicites, dont la pertinence n'est pourtant pas complètement évidente au vu de la question posée et des connaissances habituellement enseignées en probabilités élémentaires.

Cet exemple, que nous détaillons dans la section 1, montre l'intérêt d'un enseignement des probabilités axé sur la démarche de modélisation, pour interroger les liens entre situation réelle et modèle mathématique et permettre d'éclairer le choix d'un modèle plutôt qu'un autre, bien au-delà de la simple explicitation de l'univers choisi pour traduire les résultats d'une épreuve aléatoire. Cette démarche peut être effectuée à travers le cycle de modélisation de Blum et Leiß (2006), que nous décrivons dans la section 2. Dans un séminaire en ligne, Kuzniak (2022) nomme ce cycle l'*approche standard dominante* de la modélisation mathématique.

De fait, il est utilisé dans de nombreux travaux, par exemple Derouet (2022) qui détaille ses caractéristiques et ses avantages. De plus, on peut voir la présentation de la compétence « modéliser » dans le programme du cycle 4 comme une version très simplifiée, réduite à ses principales étapes, de ce cycle (Vinatier, 2025). L'une de ces étapes porte sur la validation ou l'invalidation du modèle, nous verrons plusieurs manières de la mettre en œuvre pour les modèles envisagés pour résoudre l'exercice proposé ci-dessus.

Les paradoxes que nous soulevons en section 3 traduisent pour l'essentiel le hiatus que nous venons d'évoquer entre l'importance de la connaissance de la démarche de modélisation pour la compréhension de la résolution de l'exercice et la quasi inexistence de la modélisation dans l'enseignement des mathématiques (par exemple dans les programmes du cycle 4,

[1] stephane.vinatier@unilim.fr

que nous passons rapidement en revue), et plus largement dans la culture mathématique commune. Dans la section 4, nous considérons plus particulièrement la faible place accordée à la modélisation dans l'enseignement classique des probabilités, notamment au moment de l'introduction des probabilités conditionnelles.

La section 5 est plus mathématique, elle décrit brièvement la structure d'espace probabilisé produit qui est à l'œuvre dans le « bon » modèle utilisé pour résoudre l'exercice proposé ci-dessus, ainsi que dans de nombreuses autres situations. La proposition 5 énonce une propriété de cette structure qui assure qu'elle traduit convenablement une hypothèse très courante d'indépendance de certains événements, ce qui lui donne une portée très large en tant que modèle probabiliste. Un enseignement des probabilités basé sur la démarche de modélisation pourrait difficilement s'envisager sans l'étude de cette structure.

D'autres travaux ont porté sur la démarche de modélisation en probabilité, comme Derouet (2022) qui propose *« trois catégories différentes de démarches de modélisation faisant intervenir un modèle probabiliste, pouvant être rencontrées dans l'enseignement secondaire français »* (p. 89), et qui donne de plus un état de l'art et des références utiles de ce domaine. La portée de ces travaux risque cependant d'être limitée par le manque de culture commune ou de pratique de la démarche de modélisation, que nous tentons de mettre en lumière ici, tant au niveau des élèves et des étudiants que de leurs enseignants.

Notons enfin, en passant, que le modèle mathématique utilisé pour résoudre l'exercice proposé ci-dessus a des applications pratiques en cryptologie, pour déterminer le nombre d'essais au hasard à effectuer pour trouver avec une probabilité d'au moins $\frac{1}{2}$ deux éléments ayant la même image par une *fonction de hachage*. Ce type de fonction, qui associe une *empreinte* de taille fixe (et relativement petite) aux éléments d'un très gros ensemble de « messages » possibles, est utilisée en particulier pour s'assurer de l'intégrité d'un message. Sa sécurité repose notamment sur le fait qu'il soit difficile de trouver deux messages ayant la même empreinte.

## 1. – Solution(s)… ?

Élargissons légèrement le problème soulevé par le paradoxe des anniversaires : notons $n$ le nombre de jours dans l'année, $k$ le nombre de personnes prises au hasard et calculons la probabilité de l'événement $E_k$ : « au moins deux parmi ces $k$ ont la même date d'anniversaire ».

### 1.1. - Univers ordonné

La solution classique paraît limpide au premier abord : on passe par la probabilité de l'événement contraire $\overline{E_k}$ : « toutes les personnes ont des dates d'anniversaire distinctes » et, pour déterminer sa probabilité, on compte le nombre de $k$-uplets de dates distinctes (dates à choisir parmi les $n$ possibles), c'est-à-dire le nombre d'arrangements[2] sans répétition de $k$ parmi $n$, que l'on divise par le nombre total de $k$-uplets :

$$P(\overline{E_k}) = \frac{n(n-1)\dots(n-k+1)}{n^k} = \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right) \quad (1)$$

En utilisant la majoration $1+x \leqslant e^x$ pour $x \in \mathbb{R}$, on obtient

$$P(\overline{E_k}) \leqslant exp\left(-\frac{k(k-1)}{2n}\right) \leqslant exp\left(\frac{-(k-1)^2}{2n}\right)$$

d'où, pour $p \in [0,1[$, $P(E_k) = 1 - P(\overline{E_k}) \geqslant p$ dès que $k \geq 1 + \sqrt{-2n\,ln(1-p)}$. Avec $n=365$

[2] Ici, comme souvent, le mot *arrangement* désigne une liste ordonnée, tandis que *combinaison* désignera une liste non ordonnée, dans les deux cas avec ou sans répétition autorisée.

et $p=\frac{1}{2}$, cette condition suffisante donne $k\geqslant 1+\sqrt{730\,ln(2)}\approx 23{,}494$, ce qui n'est pas tout à fait optimal : un programme calculant la valeur exacte de $P(E_k)$ à partir de l'égalité (1) donne

$$P(E_{23})\approx 0{,}507 \qquad (2)$$

donc $P(E_{23})>\frac{1}{2}$, alors que $0{,}476\approx P(E_{22})<\frac{1}{2}$.

Comme on le voit, cet exercice est très riche : dans cette solution, il fait intervenir des notions de dénombrement, de probabilités élémentaires, d'analyse, de raisonnement avec l'utilisation d'une condition suffisante, voire même de programmation pour obtenir les valeurs exactes. Il est en fait encore bien plus riche que cela, car il pose également des questions à propos de la modélisation mathématique.

### 1.2. - Modèle mathématique

Dans la solution ci-dessus, on a pris comme une évidence le fait que le modèle mathématique permettant de rendre compte de la situation « réelle » est l'espace probabilisé constitué de l'univers $\Omega=\{1,2,\dots,n\}^k$ des $k$-uplets de dates, muni de la probabilité uniforme (la tribu étant bien sûr l'ensemble des parties de $\Omega$). Cela signifie qu'on a implicitement mis un ordre sur les dates d'anniversaire, ce qui revient à dire qu'on distingue les personnes les unes des autres : chaque position dans le $k$-uplet correspond à une personne bien précise, donc on considère que le fait que $A$ soit né un 5 décembre et $B$ un 8 juillet n'est pas le même événement que si $A$ est né un 8 juillet et $B$ un 5 décembre.

Et pourtant… Quand on se pose la question de savoir si toutes les dates d'anniversaire des personnes du groupe sont distinctes, l'ordre des dates importe peu : le 5 décembre et le 8 juillet sont des dates distinctes, pourquoi s'embarrasser de savoir que l'une correspond à $A$ et l'autre à $B$ ? Ne serait-il pas ici plus naturel, pour modéliser la situation réelle, de ne pas mettre d'ordre sur les dates, c'est-à-dire de se placer dans l'univers $\Omega'$ des combinaisons, avec répétitions autorisées, de $k$ dates parmi les $n$ possibles?

### 1.3. - Univers sans ordre

Plaçons-nous donc dans $\Omega'$, que nous munissons lui aussi de la probabilité uniforme. Le cardinal de $\Omega'$ est le nombre de combinaisons, avec répétitions autorisées, de $k$ parmi $n$, égal[3] à $\binom{n+k-1}{k}$. Notons $E_k'$ le sous-ensemble de $\Omega'$ correspondant à l'événement « au moins deux personnes ont la même date d'anniversaire » et $\overline{E_k'}$ son complémentaire. Le cardinal de $\overline{E_k'}$ est le nombre de combinaisons sans répétition de $k$ dates choisies parmi les $n$ possibles, c'est-à-dire $\binom{n}{k}$, donc

$$\begin{aligned} P(\overline{E_k'}) &= \frac{\binom{n}{k}}{\binom{n+k-1}{k}} \\ &= \frac{n(n-1)\dots(n-k+1)}{(n+k-1)(n+k-2)\dots n} \\ &= \prod_{i=1}^{k-1}\frac{n-i}{n+i} \end{aligned}$$

On trouve alors, pour $n=365$, que la probabilité d'avoir au moins deux personnes avec la même date d'anniversaire parmi 23 est :

$$P(E_{23}')=1-\frac{364}{366}\times\frac{363}{367}\times\dots\times\frac{343}{387}\approx 0{,}750,$$

[3] On pourra consulter la très belle démonstration par « étoiles et barres » sur https://fr.wikipedia.org/wiki/Combinaison_avec_r%C3%A9p%C3%A9tition

ce qui est très différent de la valeur trouvée en (2) avec le modèle précédent, qui distinguait les personnes en ordonnant les dates.

### 1.4. - Conclusion ?

Quelle est « la » solution de l'exercice, quelle est la bonne valeur pour la probabilité qu'au moins deux personnes d'un groupe de 23 aient la même date d'anniversaire ? L'une des deux valeurs trouvées correspond-elle mieux à la réalité que l'autre ? Autrement dit, l'un des deux modèles est-il meilleur que l'autre ? Comment le savoir ? Pour répondre à ces questions, faisons le point sur la modélisation mathématique, en particulier sur la procédure qui va nous servir de référence.

## 2. – Modélisation mathématique

L'expression *modélisation mathématique* prend des significations très différentes selon les auteurs et les contextes, voir Kuzniak (2024) (ou Kuzniak (2022) en vidéo) pour un historique et la présentation de différents cadres théoriques. Il s'agit ici de relier la théorie mathématique des probabilités à une situation tirée du monde réel (ce sont de telles situations qui ont, historiquement, motivé le développement de la théorie). Nous avons donc besoin d'un cadre réfléchi pour régir les liens entre les mathématiques et le monde réel, ce qui est communément appelé la *modélisation extra-mathématique*.

### 2.1. - Le cycle de modélisation (selon Blum et al.)

Dans notre contexte, le *cycle de modélisation* de Blum et Leiß (2006) est le cadre de référence, utilisé par exemple dans Derouet (2022). Ce cycle se trouve déjà pour l'essentiel dans des ouvrages antérieurs, par exemple dans Blum et al. (2002, paragraphe 2.1), article préparatoire à une étude ICMI[4], plus facile à consulter en ligne, dont nous tirons les deux citations qui suivent. Ses auteurs définissent le *monde réel* de façon assez large :

> *By real world we mean everything that has to do with nature, society or culture, including everyday life as well as school and university subjects or scientific and scholarly disciplines different from mathematics*[5].

c'est-à-dire comme étant plus ou moins le complémentaire du monde mathématique[6]. Ils présentent ensuite une procédure permettant de mettre en relation ces deux mondes (*« a description of the complex interplay between the real world and mathematics »*), qu'on peut découper et schématiser comme suit (*cf.* Blum et Leiß, 2006). Partant d'une certaine situation dans le monde réel :

(1) on l'analyse pour obtenir un *modèle* de la *situation* ;

(2) on simplifie, structure, précise celui-ci, en fonction de ce qu'on en sait et de ce qu'on souhaiterait savoir ; on met ainsi au jour un *problème* (la ou les question·s qui se pose·nt), ainsi qu'un *modèle réel* de la situation ;

(3) lorsque cela est possible et semble adapté, on mathématise le modèle réel, c'est-à-dire qu'on traduit ses objets, relations, données et conditions en termes mathématiques ; en particulier, les objets

[4] International Commission on Mathematical Instruction, étude qui a donné lieu à la publication Blum et al. (2007).

[5] *Par « monde réel », nous entendons tout ce qui a trait à la nature, à la société ou à la culture, y compris la vie quotidienne ainsi que les matières scolaires et universitaires ou les disciplines scientifiques et savantes autres que les mathématiques* (notre traduction).

[6] Celui-ci n'est pas défini précisément dans l'article; on peut considérer qu'il s'agit de l'ensemble des théories construites par raisonnement déductif sur les quelques variantes des ensembles d'axiomes datant de la refondation des mathématiques au tournant du XX[e] siècle.

réels sont remplacés par des objets mathématiques et le problème posé à l'étape précédente devient une question mathématique ; on dispose alors d'un *modèle mathématique* de la situation ;

(4) on peut alors appliquer les méthodes mathématiques (propriétés, théorèmes, déductions, …) pour établir des *résultats mathématiques* portant sur les objets, relations, données, conditions du modèle mathématique, de façon à répondre à la question issue de l'étape précédente ;

(5) ces résultats mathématiques doivent ensuite être traduits dans le modèle réel, c'est-à-dire interprétés dans les termes de la situation modèle ; si une réponse à la question mathématique a été trouvée, on obtient ainsi une solution au problème de l'étape (2) ;

(6) il faut alors vérifier si la solution dans le modèle réel obtenue en interprétant les résultats mathématiques est raisonnable et appropriée par rapport aux attentes ; ceci peut conduire à *valider* le modèle mathématique ou, au contraire, à recommencer tout le processus avec un modèle modifié ou complètement différent ;

(7) enfin, une fois qu'un modèle a été validé, on le décrit précisément de manière à pouvoir le communiquer.

Ce processus est représenté dans la figure 1.

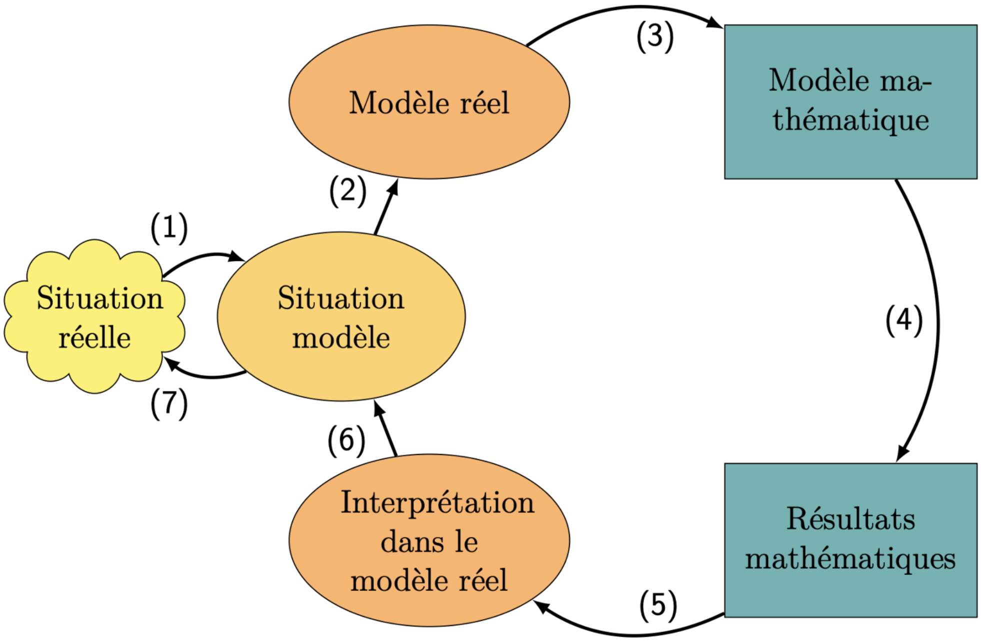


***Figure 1*** *: Cycle de modélisation mathématique, selon Blum et Leiß (2006).*

Noter que l'étape (3) consiste à établir une correspondance entre des objets réels et des objets mathématiques et que cette traduction du monde réel en langage mathématique est par essence de portée limitée (par les conditions particulières de la situation que l'on souhaite modéliser) et imparfaite, comme toute traduction. L'étape (5) utilise cette correspondance en sens inverse pour traduire les résultats portant sur les objets mathématiques en propriétés des objets réels, traduction elle aussi nécessairement imparfaite.

Tous les choix effectués aux étapes (1) à (3), voire aussi les choix de propriétés mathématiques utilisés à l'étape (4), sont des choix extra-mathématiques, qui ne sont pas justifiables par des démonstrations mathématiques. Ils relèvent de la démarche scientifique, et demandent à être validés par d'autres moyens, à l'étape (6), en confrontant les résultats obtenus aux observations déjà effectuées dans le monde réel ou à celles obtenues à l'aide de nouvelles expériences.

### 2.2. - Validation « statistique »

On a vu au point (6) du cycle de modélisation de Blum et Leiß (2006) qu'il est nécessaire, pour valider un modèle, de comparer les résultats qu'il a permis d'établir avec les observations que l'on peut faire dans le monde réel. Puisque nos modèles prévoient des taux d'apparitions de dates d'anniversaire partagées d'environ 50,7 % et 75 %, respectivement, pour les groupes de 23 personnes, on peut espérer les départager par une étude statistique portant sur des groupes de 23 personnes.

Justement, un journaliste a profité de la coupe du monde féminine de football 2023 (en Australie et Nouvelle-Zélande) pour tester le paradoxe des anniversaires sur les 32 équipes qualifiées, chacune étant composée de 23 joueuses. Ses résultats sont décrits dans un article paru sur le site web de BBC News Afrique (2023) : il a patiemment comparé les dates d'anniversaire des joueuses de chaque équipe et a constaté qu'au moins deux d'entre elles avaient la même date d'anniversaire dans 17 des 32 équipes, soit dans 53,125 % des équipes étudiées.

Cette valeur est proche de celle fournie par le premier des deux modèles utilisés ci-dessus (2), la légère différence pouvant s'expliquer par les habituelles fluctuations d'échantillonnage, au vu de la taille relativement réduite de l'échantillon sur lequel la statistique a été établie ; faible taille qui oblige aussi, malheureusement, à relativiser la portée de la validation. Il est tout de même intéressant et surprenant de constater qu'un journaliste couvrant un événement sportif a considéré de façon assez scientifique la question *a priori* très théorique du paradoxe des anniversaires ! Son enquête ne s'arrête d'ailleurs pas là : non seulement, ayant trouvé au total 24 paires de joueuses ayant des dates d'anniversaire identiques, il s'est demandé si deux (au moins) de ces dates d'anniversaire partagées étaient identiques (réponse : oui), mais il s'est aussi penché sur les correspondances partielles entre profils ADN figurant dans une base de données de l'Arizona, découvertes et comptabilisées par des scientifiques travaillant sur cette base : de telles correspondances pourraient paraître inquiétantes car susceptibles de remettre en cause des verdicts judiciaires basés sur les profils ADN ; l'auteur montre que des calculs de probabilités permettent heureusement d'évacuer ce doute.

Déjà en 2014, la BBC avait publié un article recensant les coïncidences d'anniversaire dans les équipes de football de la coupe du monde masculine (au Brésil) (Fletcher, 2014). Là aussi les résultats vont dans le sens de la confirmation de notre premier modèle : 16 équipes sur les 32 présentes à cet événement avaient au moins deux joueurs avec la même date d'anniversaire. Dans le souci d'élargir la taille de son échantillon, l'auteur avait étendu son étude aux équipes de la coupe du monde précédente (en Afrique du Sud), avec un résultat similaire : 15 équipes sur les 32 recelaient de telles coïncidences.

Au total, ces études footballistiques portent sur 96 équipes, ce qui constitue un échantillon de taille respectable. Il se peut cependant, comme cela est évoqué à la fin de Fletcher (2014), que les sportifs de haut niveau ne soient pas tout à fait représentatifs de la population générale pour ce qui est des dates d'anniversaire : les échantillons étudiés montrent une plus grande fréquence des naissances dans la première moitié de l'année chez ces joueurs, donc une répartition des dates d'anniversaire qui n'est pas totalement « au hasard ».

### 2.3. - Traduire les conditions de l'expérience aléatoire

La validation statistique *a posteriori* précédente est déjà assez convaincante, malgré l'effectif limité sur lequel les études ont porté et le léger biais constaté sur les dates d'anniversaire des sportifs de l'échantillon de (Fletcher, 2014). Un argument *a priori* va permettre de trancher définitivement entre nos deux modèles. Le point (3) de notre description du cycle de modélisation (section 2.1.) précise qu'on doit traduire les objets, relations, données *et conditions* du modèle réel en termes mathématiques. En particulier, pour traduire dans le modèle mathématique le fait que, dans le modèle réel, les personnes du groupe sont « choisies au hasard », il paraît raisonnable de souhaiter que les dates d'anniversaire des personnes du groupe soient indépendantes les unes des autres ; car autrement, les liens de dépendance entre les dates d'anniversaire, et donc entre les personnes, contrediraient le fait que le groupe est constitué au hasard.

Or le deuxième modèle ne respecte pas cette condition. Montrons-le par l'absurde, en supposant que les choix des $k$ dates sont indépendants les uns des autres. L'univers $\Omega'$ est l'ensemble des combinaisons, avec répétitions autorisées, de $k$ dates parmi les $n$ possibles. Convenons de noter une telle combinaison comme un $k$-uplet ordonné $(d_1, d_2, \ldots, d_k)$ avec $1 \leqslant d_1 \leqslant d_2 \leqslant \ldots \leqslant d_k \leqslant n$, et considérons un $k$-uplet $(d_1, d_2, \ldots, d_k)$ avec $d_1 < d_2 < \ldots < d_k$ (dates toutes distinctes). Alors l'événement élémentaire $\{(d_1, d_2, \ldots, d_k)\}$ est l'intersection des événements « une des dates est $d_1$ », « une des dates est $d_2$ », …, « une des dates est $d_k$ ». Donc, puisqu'on a supposé les choix des dates indépendants les uns des autres :

$$\begin{aligned} P\left(\{(d_1, d_2, \ldots, d_k)\}\right) &= P(\text{« une des dates est } d_1 \text{ »} \cap \ldots \cap \text{« une des dates est } d_k \text{ »}) \\ &= P(\text{« une des dates est } d_1 \text{ »}) \times \ldots \times P(\text{« une des dates est } d_k \text{ »}) \end{aligned}$$

De plus, pour tout $1 \leqslant i \leqslant k$, l'événement « une des dates est $d_i$ » est en bijection avec les listes non ordonnées, avec répétitions autorisées, de $k-1$ dates parmi les $n$ possibles, et le nombre de ces listes est $\binom{n+k-2}{k-1}$. En utilisant l'hypothèse d'équiprobabilité dans $\Omega'$, on obtient :

$$\frac{1}{|\Omega'|} = P\left(\{(d_1, d_2, \ldots, d_k)\}\right) = \left(\frac{\binom{n+k-2}{k-1}}{|\Omega'|}\right)^k,$$

d'où

$$\binom{n+k-2}{k-1}^k = |\Omega'|^{k-1} = \binom{n+k-1}{k}^{k-1},$$

ce qui est clairement faux en général.

#### *Conclusion*

Si on se place dans l'univers $\Omega'$ des listes non ordonnées de $k$ parmi $n$ avec répétitions autorisées, on ne peut pas à la fois faire l'hypothèse d'équiprobabilité (des événements élémentaires) et imposer la condition « réaliste » que les choix des différentes dates sont indépendants. Notre deuxième modèle, qui faisait l'hypothèse d'équiprobabilité, ne respecte pas la condition d'indépendance des dates, il ne rend donc pas bien compte de la situation réelle, si bien que nous devons nous résoudre à l'abandonner.

On verra dans la section 5 que l'indépendance des dates de naissance des différentes personnes est automatiquement vérifiée dans le premier modèle, du fait même du procédé de construction de l'espace probabilisé utilisé, ce qui nous permet de valider ce modèle mathématique également du point de vue de la prise en compte des conditions du modèle réel.

### *Limites*

Notons tout de même qu'en tant que modèle mathématique appliqué à une situation réelle, il a nécessairement des limites dans ses capacités d'explication et de prédiction et que le passage au monde réel rend vaine toute tentative de démonstration mathématique de sa pertinence : dès lors qu'on fait un lien entre les mathématiques et le monde réel, que par suite on sort du monde mathématique, le raisonnement axiomatico-déductif n'est plus applicable et doit laisser la place à d'autres critères de validation, comme ceux qu'on vient de mettre en œuvre, qui ne peuvent par nature jamais être définitifs. De fait, la démarche scientifique consiste pour une bonne part à imaginer des modèles (mathématiques ou autres) des phénomènes étudiés et à en déterminer la pertinence, en particulier les limites et conditions d'application. Tout modèle, aussi séduisant soit-il, ne décrit convenablement la réalité que dans certaines limites[7].

## 3. – Paradoxes

Si le paradoxe des anniversaires est dénommé ainsi parce qu'il est surprenant, à première vue, qu'il suffise de 23 personnes choisies au hasard pour avoir plus d'une chance sur deux qu'au moins deux d'entre elles aient la même date d'anniversaire, l'étude de la résolution de cet exercice est révélatrice de plusieurs paradoxes.

### *Avec ou sans ordre*

Le premier paradoxe que nous venons de mettre au jour est le suivant : alors même que la question posée ne dépend en aucune façon du fait que les dates d'anniversaire soient ordonnées ou non, c'est-à-dire que l'on distingue ou non les personnes, on est amené à choisir un modèle avec un ordre sur les dates pour respecter au mieux les conditions du modèle réel de la situation. L'alternative consistant à abandonner l'hypothèse d'équiprobabilité dans le modèle sans ordre rendrait certainement les calculs et le raisonnement mathématiques nettement plus complexes et, de toute façon, on verra à la section 5 que les propriétés des espaces probabilisés produits rendent leur usage à la fois naturel et en quelque sorte universel pour assurer l'indépendance de certains types d'événements, afin de traduire une condition du modèle réel. Ce qui élucidera complètement ce premier « paradoxe » et pourrait par ailleurs justifier que la structure d'espace probabilisé produit soit plus largement enseignée, en tant qu'outil permettant la modélisation de très nombreuses situations (penser en particulier au schéma de Bernoulli !).

### *Structure d'espace probabilisé produit*

Le fait qu'elle soit très peu présente dans les cours de probabilités, quel que soit le niveau, pourrait être le deuxième paradoxe révélé par notre étude : alors que l'enseignement des probabilités s'est beaucoup développé dans le supérieur et le secondaire (et désormais également dans le primaire), le plus souvent en faisant le lien avec des situations réelles, il est paradoxal que les programmes d'enseignement ne donnent pas aux apprenants les connaissances mathématiques leur permettant de traduire certaines hypothèses du modèle réel et, dans le cas du paradoxe des anniversaires, de comprendre le choix du modèle. On détaille les connaissances relatives à la structure d'espace probabilisé produit à la section 5, où l'on

[7] Il n'est pas inintéressant de constater, à ce titre, que le modèle de la Terre plate est tout à fait valide pour de très nombreuses applications, par exemple pour mesurer la longueur de la plupart de nos déplacements ou encore les aires des parcelles agricoles ou urbaines. Une meilleure connaissance de la démarche de modélisation aiderait peut-être à dépasser le clivage parfois stérile entre les « platistes » et les autres.

verra qu'elles sont tout à fait à la portée des étudiants en licence de mathématiques, preuves comprises, et même quasiment accessibles aux élèves de première et terminale en spécialité mathématiques, puisque le produit cartésien d'ensembles et la notion d'indépendance sont à leur programme ; il ne leur manque que la formalisation de ce qu'est une probabilité (ce qui paraît étonnant à ce niveau d'étude).

### *Enseignement de la modélisation*

Bien sûr, l'outil essentiel pour donner du sens à l'application des probabilités à des situations réelles est le cycle de modélisation de Blum et Leiß, ou un équivalent, dans une version simplifiée et adaptée au niveau des élèves si nécessaire. Le troisième paradoxe révélé par le paradoxe des anniversaires est donc celui-ci : il est essentiel, pour comprendre l'application des probabilités à des situations réelles, d'avoir en tête les principales étapes de la démarche de modélisation mathématique, et pourtant celles-ci ne sont en pratique jamais clairement présentées aux apprenants.

On les voit pourtant apparaître, sous une forme très simplifiée, au niveau du préambule du programme du cycle 4, dans la description de la compétence « Modéliser » (MENJS, 2020, p. 129–130) : *« Traduire en langage mathématique une situation réelle* […] *Valider ou invalider un modèle »*, ce qu'on peut interpréter comme résumant les étapes (3)[8] et (5)-(6) du cycle de modélisation, respectivement ; la phrase placée juste avant : *« Reconnaître un modèle mathématique (proportionnalité, équiprobabilité) et raisonner dans le cadre de ce modèle pour résoudre un problème »* pouvant être vue comme une adaptation à certains outils mathématiques du collège des étapes (3)-(4). Malheureusement, ces bonnes intentions ne sont quasiment pas reprises dans les contenus du programme et, lorsqu'elles le sont, c'est sans aucune mention des modalités de leur mise en œuvre, du cycle de modélisation ou de ses principales étapes.

Même chose, hélas, dans la ressource d'accompagnement MEN (2016) qui, si elle décrit succinctement ces étapes et pose clairement la distinction entre modèle et réalité, n'explique aucunement comment présenter la démarche aux élèves, pas plus qu'elle n'en fait un attendu de l'enseignement. De fait, ce document parle plus de l'évolution des connaissances mathématiques des élèves entre le cycle 2 et le cycle 4 que de modélisation et, lorsqu'il aborde la mise en œuvre de celle-ci, il insiste surtout sur les inconvénients des activités de modélisation, en terme de temps nécessaire (pour accomplir tout le processus sur un exemple) et de portée pour les apprenants (en cas de travail sous forme de projet). Ses auteurs ne semblent considérer la démarche de modélisation que comme un type d'activité, et non en tant que démarche à enseigner et compétence à acquérir[9]. Leur cas n'est pas isolé : dans la littérature, de nombreux articles présentent la modélisation comme une activité complexe à mettre en œuvre, sujet par ailleurs de riches expérimentations et réflexions, et non comme le cadre naturel de toute application des mathématiques au monde réel. Pourtant, pour développer la compétence « Modéliser » telle qu'elle est décrite dans le préambule de MENJS (2020), ne faudrait-il pas introduire les principales étapes de la modélisation mathématique sur des exemples très simples ?

Il est à craindre que les obstacles que nous venons de citer, couplés à l'absence de réelle incitation, aient pour effet que la modélisation

[8] ou (2)-(3) voire (1)-(3), selon d'où l'on part.

[9] Pour ajouter à la confusion, les deux premières références bibliographiques citées dans MEN (2016) sont essentiellement consacrées à un tout autre type de modélisation, parfois dite intramathématique, qui ne gère pas du tout les rapports entre les mathématiques et le monde réel.

mathématique ne soit que très peu enseignée au cycle 4.

### *Donner plus de sens aux applications des mathématiques*

Les exercices du niveau collège dans lesquels la mise en œuvre de cette démarche serait simple et éclairante ne manquent pas. Le prix des pommes est-il proportionnel à leur quantité ? Cette question qui mélange monde réel et mathématique ne prend tout son sens que si on la considère à travers la démarche de modélisation, et devient alors : dans quelles limites (ou conditions) le modèle linéaire avec un certain cœfficient de proportionnalité est-il valide pour représenter le prix des pommes en fonction de leur quantité ? Amener les élèves sur ce chemin ne demande que de leur fournir un cadre leur permettant de poser clairement la distinction entre le modèle réel et le modèle mathématique (le cycle de modélisation simplifié, comme cela est suggéré dans MENJS (2020)), tout en développant une compétence très utile pour la suite de leur parcours. Le groupe *Modélisation au cycle 4* de L'IREM de Limoges a conçu et testé en classe une fiche d'activité sur ce thème. Ce travail a été présenté lors d'un atelier au colloque CORFEM 2024, qui devrait prochainement faire l'objet d'un compte-rendu dans les Actes de ce colloque (Vinatier, 2025).

Enseigner le cycle de modélisation sous une forme simplifiée serait certainement porteur de beaucoup de sens pour les mathématiques du collège en général, ainsi que pour leur application aux autres siences, et permettrait aux élèves, au cours de leur apprentissage des probabilités, d'avoir la familiarité nécessaire avec la démarche de modélisation pour comprendre au mieux les liens entre cette nouvelle théorie mathématique et le monde réel. À défaut, les quelques mentions concernant la modélisation dans la partie Statistiques et probabilités du programme de seconde générale et technologique MEN (2019a) ont de fortes chances de rester lettre morte.

## 4. – Modélisation et enseignement des probabilités

Beaucoup de problèmes posés aux élèves ou aux étudiants, en probabilités, sont déjà sous la forme de modèles réels (la simplification et la structuration de la situation réelle sont déjà effectuées, le problème clairement dégagé). Il s'agit donc, pour donner leur vrai sens aux calculs et déductions mathématiques effectués à l'étape (4), d'expliciter les étapes (3) — mise en correspondance, aussi fidèle que possible, des données et conditions avec les objets et relations d'un modèle mathématique, (5) — traduction des résultats mathématiques dans le modèle réel, en utilisant la correspondance inverse, et (6) — validation ou non du modèle choisi, en fonction de la pertinence des résultats obtenus.

### *Univers*

Lors des premiers temps de l'enseignement des probabilités, on met souvent l'accent sur la description de l'univers (en général muni de la tribu de toutes les parties), voire de la probabilité utilisée (souvent uniforme). On peut voir là une esquisse de mise en œuvre du cycle de modélisation simplifié que nous venons d'ébaucher, c'est-à-dire les étapes (3) à (6) du cycle de la section 2.1, avec de fortes limitations :

- bien souvent seules les étapes (3) et (4) sont concernées, parfois l'étape (5) si on traduit le résultat mathématique dans les termes de la situation d'origine, rarement la (6) ;
- on n'est pas toujours conscient des conditions que le modèle choisi va satisfaire, comme on l'a vu ci-dessus : les propriétés des différents modèles disponibles

n'ont pas été établies au préalable pour permettre un choix éclairé du modèle ;
- le cycle de modélisation simplifié n'a en général pas été introduit auparavant, indépendamment de son utilisation pour appliquer les probabilités, ce qui ne permet pas aux apprenants d'en saisir le sens et l'intérêt.

De plus, la description de l'univers est en général abandonnée dès lors qu'on utilise les probabilités conditionnelles (et encore plus avec les variables aléatoires).

### *Un exercice typique*

Ainsi, l'exercice

> *Une urne contient 5 boules noires et 10 boules blanches. On tire deux boules sans remise. Quelle est la probabilité d'obtenir une boule blanche au deuxième tirage ?*

est habituellement résolu en appliquant la formule des probabilités totales (ou en utilisant l'arbre pondéré qui décrit la situation), sans se soucier de savoir dans quel univers on travaille :

$$\begin{aligned} P(B_2) &= P(B_2|N_1)P(N_1)+P(B_2|B_1)P(B_1) \\ &= \frac{10}{14}\times\frac{1}{3}+\frac{9}{14}\times\frac{2}{3} \\ &= \frac{2}{3} \end{aligned}$$

où, par exemple, $B_2$ est l'événement « obtenir une boule blanche au 2$^{e}$ tirage ». D'ailleurs, l'univers dans lequel on travaille implicitement pour calculer les probabilités des différentes branches de l'arbre correspond au résultat du tirage d'une boule (éventuellement restreint par une condition), et non à celui de l'épreuve aléatoire décrite dans l'énoncé, dont le résultat est un *couple* de boules distinctes.

Notons que, dans ce cas, il existe une solution élémentaire plus simple : travaillons-dans l'univers $\Omega$ composé des listes ordonnées sans répétition à 2 éléments parmi 15, de cardinal $|\Omega|=\frac{15!}{13!}=15\times14=210$ ; le nombre d'issues pour lesquelles l'événement $B_2$ est réalisé, c'est-à-dire de couples de $\Omega$ dont la deuxième composante correspond à une boule blanche, est $10\times14$ (en considérant d'abord la deuxième composante du couple), donc

$$P(B_2)=\frac{10\times14}{15\times14}=\frac{2}{3}$$

Il apparaît que, dans sa solution classique, cet exercice n'est pas considéré comme une activité de modélisation d'une situation réelle, mais plutôt comme un habillage rudimentaire de la formule des probabilités totales, à appliquer de façon quasi mécanique aux données numériques. Le fait que l'univers dans lequel on travaille, lorsqu'on utilise les probabilités conditionnelles, n'est pas composé d'objets mathématiques codant les résultats de l'épreuve aléatoire « réelle » (mais uniquement ceux du premier tirage ici) est passé sous silence : on abandonne la description de l'univers, c'est-à-dire le lien entre modèle réel et modèle mathématique, plutôt que d'expliciter ce changement de point de vue.

### *Flou artistique*

Le lien n'étant plus explicité entre la théorie mathématique et la situation à laquelle on l'applique, on peut se demander quel sens les apprenants donnent à cette pratique et aux résultats qu'ils obtiennent. L'absence d'explicitation de la démarche de modélisation peut d'une part donner à la théorie des probabilités un caractère flou et indistinct, celui d'une discipline à la fois mathématique et en dehors des mathématiques, sans qu'on sache jamais vraiment de quel côté de la frontière on se trouve ; elle risque aussi, d'autre part, de faire penser que le résultat mathématique est une vérité démontrée du monde réel, faute de distinction clairement posée et de cadre pour réfléchir le lien entre les deux. Cette confusion est tellement généralisée qu'une très grande partie du vocabulaire et des expressions utilisés en probabilités en portent la marque et la renforcent, nous y revenons ci-dessous.

Même en début d'apprentissage, lorsqu'on prend encore soin de déterminer l'univers adéquat au début de chaque exercice, quel sens cela a-t-il si on n'inscrit pas cette étape dans le cadre de la démarche de modélisation, sinon celui d'une sorte de rituel propre à cette discipline hybride ? Qu'on pourra ensuite abandonner sans remords, ce qu'on fait souvent assez vite. On l'a vu avec l'exercice typique sur la formule des probabilités totales ci-dessus. On le voit également dans la présentation des lois usuelles (notamment binomiale), qui ne fait le plus souvent aucune mention de l'univers sous-jacent, introduisant des variables aléatoires sans indiquer l'ensemble de départ de ces applications.

### *L'importance du vocabulaire et des expressions*

Les termes qu'on utilise en probabilités portent la marque d'une forte imprégnation de l'absence de distinction entre monde réel et théorie mathématique. Ainsi, le même mot « événement » est habituellement utilisé pour désigner à la fois certains types de résultats de l'épreuve aléatoire (côté modèle réel) et les sous-ensembles correspondants de l'univers (côté modèle mathématique), comme on peut le constater rétrospectivement dans les solutions proposées ci-dessus au paradoxe des anniversaires.

Dans le même ordre idée, le concept d'*épreuve aléatoire* appartient clairement au modèle réel, mais pas au modèle mathématique. En donner une « définition » comme on le fait pour les objets mathématiques est susceptible de brouiller la frontière entre les deux modèles et d'entretenir la confusion entre ce qui relève de l'un et de l'autre. Bien sûr, le mot « définition » a en français une acception assez large pour permettre un tel usage ; cependant, dans un contexte mathématique, il pourrait être bénéfique de le réserver à la définition rigoureuse des concepts mathématiques, pour renforcer la différenciation entre le monde réel et le monde mathématique[10].

Enfin, la manière de rédiger les énoncés des exercices de probabilités gomme à l'avance la distinction entre théorie mathématique et monde réel. Dans notre principal exemple, la probabilité de l'événement « au moins deux personnes dans le groupe ont la même date d'anniversaire » est évoquée dans l'énoncé comme si elle existait par elle-même, dans la réalité, sans qu'il soit nécessaire de la construire (à travers la démarche de modélisation, c'est-à-dire ici l'établissement d'un univers et d'une probabilité), occultant *de facto* la nécessité (et la possibilité) d'effectuer des choix. Il en va de même dans la plupart des exercices (peut-être tous ?), comme celui sur la formule des probabilités totales ci-dessus. Certes un énoncé respectant la volonté d'inscrire les probabilités dans la démarche de modélisation serait sans doute plus long à écrire, mais est-ce si grave, si cela pouvait l'enrichir d'une dimension très significative et si l'application des probabilités aux situations réelles y gagnait du sens, à la fois pour les apprenants et leurs enseignants ?

La question qui se pose est de traduire la probabilité (ici $\frac{1}{2}$) dans le modèle réel, où les probabilités n'existent pas. Peut-être par le biais de la loi des grands nombres, de la façon suivante :

[10] La distinction fondamentale entre « définition » et « proposition » n'est pas toujours claire dans l'esprit des élèves, des étudiants et même de certains rédacteurs de manuels, ce qui fait une autre bonne raison de réserver le mot « définition » à son usage principal en mathématiques.

*Trouver le plus petit entier k tel que, en considérant un nombre suffisamment grand de groupes de k personnes choisies au hasard, il y en ait plus d'un sur deux qui contienne au moins deux personnes avec la même date d'anniversaire.*

## 5. – Espace probabilisé produit

Nous avons évoqué ci-dessus la structure d'*espace probabilisé* produit comme outil indispensable à la modélisation de certaines situations réelles, pour traduire des conditions d'indépendance de certains événements. Cette structure est peu présente dans les cours de probabilités en tant que telle, nous en donnons ci-dessous une présentation rapide, avec application directe à l'exemple du paradoxe des anniversaires.

Dans les programmes de l'enseignement secondaire, la notion de probabilité n'est pas formalisée, même dans le cas d'un univers fini (où c'est simplement une fonction de l'ensemble des parties de l'univers vers l'intervalle $[0,1]$, envoyant l'univers sur 1, et *additive*). Le contraire serait étonnant puisque la notion de fonction elle-même n'est présente quasiment que dans le cas particulier des fonctions de la variable réelle (voire aussi des fonctions sur $\mathbb{N}$ pour introduire les suites en seconde), sauf dans les parties « algorithmique » où d'autres fonctions sont envisagées. On peut regretter que les programmes donnent une vision aussi restrictive de cet objet central des mathématiques et de leurs applications, et ne profitent pas de la notion de probabilité pour élargir le point de vue (en même temps que celui sur les ensembles, puisqu'il faudrait aussi définir l'ensemble des parties d'un ensemble).

De ce fait, la probabilité produit n'est pas formellement définie en tant qu'objet mathématique (comme dans la proposition-définition 1 ci-dessous), même dans le programme de spécialité de mathématiques de terminale générale (MEN, 2019b) où elle apparaît pour modéliser une succession d'épreuves indépendantes. Plus généralement, « la » probabilité qui est utilisée dans l'enseignement secondaire semble être une donnée (du cours, du problème, de l'exercice) bien plus qu'un choix à faire au cours de la modélisation.

À travers ce prisme, il est difficile de distinguer l'intérêt de la construction de l'espace probabilisé produit : d'une part il s'agit du seul choix disponible, en spécialité mathématiques de la classe de terminale, pour modéliser une succession d'épreuves indépendantes, d'autre part la propriété essentielle que certains types d'événements sont automatiquement indépendants dans cette structure (la proposition 5 ci-dessous) n'est pas mentionnée. Ici encore, le programme ne se soucie guère des enjeux de la modélisation, pour se concentrer sur l'application non réfléchie d'une technique imposée et admise.

### 5.1. - Définition

On ne présente ici que le cas le plus simple des espaces probabilisés finis. On ne mentionne pas les tribus associées qui sont tout simplement leurs ensembles de parties. On vérifie facilement la propriété contenue dans l'énoncé suivant.

***Proposition-définition 1***

Soient $(\Omega_1, P_1)$ et $(\Omega_2, P_2)$ des espaces probabilisés finis. Notons $\Omega = \Omega_1 \times \Omega_2$ et $\mathcal{P}(\Omega)$ l'ensemble de ses parties. L'application $P : \mathcal{P}(\Omega) \to [0,1]$ définie par $P(\varnothing)=0$, $P(\{(a,b)\}) = P_1(\{a\})P_2(\{b\})$ pour tous $a \in \Omega_1$, $b \in \Omega_2$ et par additivité[11] sur $\Omega$ tout entier, est une probabilité sur $\Omega$. L'espace proba-

[11] C'est-à-dire que l'image d'une union disjointe (dénombrable) de parties est la somme des images des parties. Puisqu'on est dans le cas fini, toute partie $E$ de $\Omega$ est union finie des singletons $\{x\}$ avec $x \in E$.

bilisé $(\Omega, P)$ est appelé *espace probabilisé produit* de $(\Omega_1, P_1)$ et $(\Omega_2, P_2)$.

On généralise sans problème la construction au produit de $n \geqslant 2$ espaces probabilisés. On vérifie que si chacun des univers $\Omega_i$, $1 \leqslant i \leqslant n$, est muni de la probabilité uniforme, l'espace produit l'est également.

***Exemple 2***

Pour modéliser le choix de 23 dates d'anniversaire parmi les 365 jours de l'année, on considère les univers

$$\Omega_1 = \Omega_2 = \dots = \Omega_{23} = \{1, 2, \dots, 365\}$$

tous munis de la probabilité uniforme $P_1 = P_2 = \dots = P_{23}$ vérifiant $P_i(\{d\}) = \frac{1}{365}$ pour tout entier $d$ avec $1 \leqslant d \leqslant 365$.

L'univers produit est le produit cartésien $\{1, 2, \dots, 365\}^{23}$, ensemble des 23-uplets d'éléments de $\{1, 2, \dots, 365\}$, son cardinal est $365^{23}$ et la probabilité produit d'un événement élémentaire $\{(d_1, d_2, \dots, d_{23})\}$ est

$$\begin{aligned} & P(\{(d_1, d_2, \dots, d_{23})\}) \\ = \; & P_1(\{d_1\}) P_2(\{d_2\}) \dots P_{23}(\{d_{23}\}) \\ = \; & \left(\frac{1}{365}\right)^{23} \\ = \; & \frac{1}{365^{23}}. \end{aligned}$$

C'est la probabilité uniforme sur l'espace produit.

### 5.2. - Indépendance des événements propres

Soit $(\Omega, P)$ l'espace probabilisé produit de $n$ espaces probabilisés $(\Omega_1, P_1)$, $(\Omega_2, P_2)$, …, $(\Omega_n, P_n)$. Les *pavés* de $\Omega$ sont les événements $E \subset \Omega$ de la forme $E = A_1 \times A_2 \times \dots \times A_n$ où $A_i \subset \Omega_i$ pour tout entier $i$ tel que $1 \leqslant i \leqslant n$.

***Proposition 3***

Soit $E = A_1 \times \dots \times A_n$ un pavé de $\Omega$, alors

$$P(E) = P_1(A_1) \dots P_n(A_n).$$

*Démonstration*

On a $E = \{(a_1, \dots, a_n),\ a_i \in A_i\}$, donc, par additivité de $P$ :

$$\begin{aligned} P(E) &= \sum_{a_1 \in A_1} \dots \sum_{a_n \in A_n} P(\{(a_1, \dots, a_n)\}) \\ &= \sum_{a_1 \in A_1} \dots \sum_{a_n \in A_n} P_1(\{a_1\}) \dots P_n(\{a_n\}) \\ &= \left(\sum_{a_1 \in A_1} P_1(\{a_1\})\right) \dots \left(\sum_{a_n \in A_n} P_n(\{a_n\})\right) \\ &= P_1(A_1) \dots P_n(A_n). \end{aligned}$$

***Définition 4***

Un événement $E \subset \Omega$ est *propre* à $\Omega_i$ s'il existe $A_i \subset \Omega_i$ tel que

$$E = \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n.$$

Cela signifie que l'événement $E$ n'impose une contrainte que dans l'univers $\Omega_i$. On note que les événements propres sont des pavés particuliers, on en déduit facilement ce qui suit.

***Proposition 5***

Soient $E_1, \dots, E_n \subset \Omega$ tels que, pour chaque $i$ dans $\{1, \dots, n\}$, $E_i$ soit un événement propre à $\Omega_i$. Alors les événements $E_i$, $1 \leqslant i \leqslant n$, sont deux à deux indépendants pour la probabilité produit.

***Exemple 6***

On revient au choix de 23 dates d'anniversaire parmi les 365 jours de l'année, avec les mêmes notations que ci-dessus. On fixe un 23-uplet de dates $(d_1, d_2, \dots, d_{23}) \in \Omega$ et, pour chaque indice $i$ tel que $1 \leqslant i \leqslant 23$, on considère l'événement $E_i$ : « la $i$-ème date choisie est $d_i$ ». On a donc

$$E_i = \Omega_1 \times \ldots \times \Omega_{i-1} \times \{d_i\} \times \Omega_{i+1} \times \ldots \times \Omega_{23},$$

autrement dit $E_i$ est propre à $\Omega_i$. Par la proposition 5, il s'ensuit que les $E_i$ sont par construction deux à deux indépendants, c'est-à-dire que les choix des différentes dates sont indépendants les uns des autres dans ce modèle.

## Conclusion

Le paradoxe des anniversaires nous a permis d'explorer d'assez près la frontière souvent floue, dans le domaine des probabilités, entre solution mathématique et situation réelle. Nous l'avons fait avec l'aide du très standard cycle de modélisation de Blum et Leiß, qui permet justement de réfléchir à la manière d'utiliser des théories mathématiques, ici probabilistes, pour répondre à une question posée dans le monde réel, et en particulier d'interroger la validité des modèles choisis, en fonction de la situation à laquelle on les applique. Nous avons constaté que cette démarche, réduite éventuellement à ses principales étapes, est essentielle pour comprendre l'enseignement des probabilités élémentaires, du fait qu'on les présente — à juste titre — en lien avec des situations réelles (jeux de hasard par exemple).

Ce constat nous amène à regretter que la modélisation mathématique ne soit pas un objet d'enseignement des programmes de mathématiques dès le cycle 4, bien que ces programmes mettent en avant la compétence « Modéliser » et qu'on puisse interpréter la présentation qu'ils en font comme une version très simplifiée du cycle de Blum et Leiß. Il nous semble qu'une initiation des élèves aux principales étapes de cette démarche les aiderait à mieux comprendre le sens des solutions mathématiques des exercices faisant intervenir les probabilités, voire de tous les exercices de mathématiques basés sur une situation réelle.

Plus largement, il nous semble que la démarche de modélisation mathématique devrait devenir partie intégrante de la culture commune des mathématiciens, qu'ils soient universitaires ou enseignants du secondaire, afin que sa mise en œuvre au sein de l'enseignement des probabilités paraisse parfaitement naturelle. Cela nécessiterait sans doute un gros effort de communication ou de formation, dans la mesure où le vocabulaire même des probabilités élémentaires et la manière dont sont rédigés les exercices montrent au contraire à quel point théorie mathématique et situation réelle sont mélangées dans leur enseignement traditionnel.

La réflexion sur le paradoxe des anniversaires a aussi fait émerger la structure d'espace probabilisé produit, comme modèle universel pour assurer l'indépendance de certains types d'événements. Un enseignement des probabilités soucieux de prendre en compte la démarche de modélisation devrait mettre l'accent sur la propriété intrinsèque de cette structure, qui justifie son emploi dans de nombreuses situations, à commencer par le schéma de Bernoulli.

Enfin, nous n'avons abordé qu'à la marge les probabilités conditionnelles et pas du tout la présentation des probabilités en termes de variables aléatoires. L'inscription de ces usages très répandus de la théorie des probabilités dans le cycle de modélisation n'est pas souvent effectuée, alors qu'elle n'est peut-être pas toujours évidente, comme l'a montré l'exercice « typique » d'application de la formule des probabilités totales. Il y a peut-être là tout un champ d'investigations complémentaires à mener.

**Stéphane VINATIER**
IREM de Limoges
XLIM UMR 7252 CNRS - Université de Limoges

## Références bibliographiques

Blum, W. et al. (2002). ICMI Study 14 : Applications and modelling in mathematics education - Discussion document. *Educational Studies in Mathematics, 51*, 149–171.
https://doi.org/10.1023/A:1022435827400

Blum, W. & Leiß, D. (2006). Filling up - The problem of independence-preserving teacher interventions in lessons with demanding modelling tasks. Dans M. Bosch (dir.), *Proceedings of the 4th Conference of the European Society for Research in Mathematics Education (2005) (p. 1623–1633)*. Europeans research in mathematics education 4. CERME.

Blum, W., Galbraith, P. L., Henn, H.-W. & Niss, M. (dir.) (2007). *Modelling and applications in mathematics education*. New ICMI study series. Springer.

Derouet, C. (2022). Caractérisation de démarches de modélisation probabiliste. *Annales de Didactiques et de Sciences Cognitives, 27*, 89–131.

Fletcher, J. (2014). The birthday paradox at the World Cup. *BBC News (16 juin 2014)*.
https://www.bbc.com/news/magazine-27835311.

Kuzniak, A. (2022). *Enseigner la modélisation mathématique pour enseigner les mathématiques : une dynamique problématique*. Séminaire de l'IREMS de Paris, 16 mars 2022.
https://video.irem.univ-paris-diderot.fr/w/sGUYR3ZyQeiv4FpBGa3EGq

Kuzniak, A. (2024). Enseignement de la modélisation mathématique et construction du travail mathématique : une dynamique problématique. *Recherches en Didactique des Mathématiques. Synthèses et perspectives en didactique des mathématiques. Preuve, modélisation et technologies numériques (Numéro spécial)*, 91–110.

Vinatier, S. (2025). Scénario d'enseignement et de formation s'appuyant sur la notion de cycle de modélisation. *Soumis pour publication*.

BBC News Afrique (2023). Le problème mathématique inattendu à l'œuvre pendant la Coupe du monde de football féminin. *BBC News Afrique (21 octobre 2023)*
https://www.bbc.com/afrique/monde-66695999

### Références institutionnelles

MEN (2016). Modéliser. *Ressources d'accompagnement du programme de mathématiques au cycle 4*. Éduscol.
https://eduscol.education.fr/document/17218/download.

MEN (2019a). Programme de mathématiques de seconde générale et technologique. *Bulletin officiel de l'Éducation nationale 1 (22 janvier 2019)*.
https://eduscol.education.fr/document/24553/download.

MEN (2019b). Programme de spécialité de mathématiques de terminale générale. *Bulletin officiel de l'Éducation nationale 8 (25 juillet 2019)*.
https://eduscol.education.fr/document/24568/download.

MENJS (2020). Programmes du cycle 4. En vigueur à la rentrée 2020. *Bulletin officiel de l'Éducation nationale, de la Jeunesse et des Sports 31 (30 juillet 2020)*.

https://eduscol.education.fr/document/621/download.