Une enseignante de mathématiques profite d'un moment de pause dans son cours pour souligner l'anniversaire de deux élèves. Bien évidemment, les élèves saisissent l'occasion pour offrir leur souhait d'anniversaire d'une façon plus ou moins synchronisée à leurs collègues de classe. Au travers de toute cette joyeuse cacophonie, un commentaire retient l'attention de l'enseignante : « Deux élèves, dans la même classe, ont la même date d'anniversaire ! Quelles sont les chances que cela se produise ? ». L'enseignante attrape la balle au bond et réplique à toute la classe : « Mathématiquement parlant, c'est assez probable ! ».

Véronique Boutet
Collège Durocher
St-Lambert

L'enseignante invite la classe à discuter de cette probabilité. Une première idée proposée est d'énumérer tous les jours de l'année et de regarder la probabilité que deux élèves aient ce jour comme anniversaire. Cela sera difficile. Une autre approche est proposée : s'intéresser aux cas où deux élèves partagent la même date d'anniversaire, ou encore trois élèves, quatre élèves, cinq élèves ... Un cas n'exclut pas l'autre ; dans la même classe, il peut y avoir deux élèves qui partagent leur anniversaire ainsi que trois autres élèves qui partagent eux aussi la leur. Le nombre de cas possibles est donc assez grand.

« Madame, c'est beaucoup trop long ! »

La classe convient rapidement qu'il sera plus facile de réfléchir à la situation en calculant plutôt la probabilité que tous les élèves aient une date d'anniversaire différente. En effet, les deux événements suivants s'excluent mutuellement :

« Tous les élèves ont une date d'anniversaire différente. »

« Au moins deux élèves partagent la même date d'anniversaire. »

En tout temps, exactement un des événements est vrai. On dit alors qu'ils sont *complémentaires* ; la somme de leur probabilité est 1.

Le calcul sera effectué en considérant qu'une année possède 365 jours. On va compter, d'une part, le nombre de combinaisons possibles des dates d'anniversaire dans un groupe de 30 élèves et, d'autre part, le nombre de celles composées cette fois-ci uniquement de dates d'anniversaire distinctes.

Pour le premier cas, on a 365 possibilités pour la date d'anniversaire du premier élève et 365 possibilités pour celle du deuxième élève. Il y a donc $365 \times 365 = 365^2$ combinaisons possibles pour les dates d'anniversaire de deux élèves. En généralisant pour 30 élèves, on obtient 365^{30} combinaisons possibles.

Pour le deuxième cas, il faut répondre à la question suivante : parmi ces combinaisons, combien sont composées de dates d'anniversaire uniques ? Commençons en considérant que n'importe quel jour peut être la date d'anniversaire du premier élève. Par la suite, puisque nous ne voulons pas que deux élèves partagent la même date d'anniversaire, celle du deuxième élève se trouve parmi les 364 jours restants. La date d'anniversaire du troisième élève est différente de celle des deux

premiers, donc parmi les 363 jours restants, et ainsi de suite pour l'ensemble des élèves de la classe. On a donc $365 \times 364 \times \cdots \times 337 \times 336$ combinaisons possibles de dates d'anniversaire.

La probabilité que tous les élèves aient des dates d'anniversaire distinctes est donc :

$$\frac{365 \times 364 \times \cdots \times 337 \times 336}{365^{30}}.$$

Finalement, en utilisant le fait que les deux événements mentionnés plus haut sont *complémentaires*, la probabilité cherchée, soit celle qu'au moins deux élèves partagent leur anniversaire, est

$$1 - \left(\frac{365 \times 364 \times \cdots \times 337 \times 336}{365^{30}}\right)$$
$$\approx 0{,}7063 = 70{,}63\%.$$

Surprise ! 70 % des chances, c'est beaucoup ! Il s'avère donc qu'il est fort probable que deux élèves aient la même date d'anniversaire. Affaire conclue !

« Madame, votre calcul n'est pas bon, ce n'est pas la vraie vie ! »

L'enseignante comprend bien l'intervention de l'élève. Le raisonnement mathématique derrière ce calcul de probabilité simplifie grandement la réalité. Entre autres, il ne tient pas compte des années bissextiles et suppose qu'il y a autant de naissances chaque jour de l'année, donc que les jours sont *équiprobables*. La classe accepte de passer à autre chose, mais l'enseignante garde cela en tête. À son tour de se poser la question : quelles sont les chances ? Les « vraies » chances ! Serait-ce possible de calculer la probabilité qu'au moins deux élèves partagent la même date d'anniversaire tout en tenant compte de ces nouveaux éléments ? Non ! Trop de facteurs influencent les naissances pour être en mesure de calculer cette probabilité de façon précise. Cependant, il n'est pas nécessaire d'avoir une réponse parfaitement précise. Une estimation basée sur des données réelles pourrait suffire à vérifier si les simplifications faites dans les calculs ci-dessus faussent significativement la probabilité finale.

Pour ce faire, l'enseignante avait besoin d'avoir accès à des données des naissances quotidiennes au Québec, ce qui lui fut gracieusement fourni par l'Institut de la statistique du Québec. Il va sans dire que l'enseignante tient à les remercier chaleureusement sans quoi elle n'aurait pu faire ses estimations. En connaissant le nombre de naissances par jour au Québec, il est possible d'estimer la probabilité de naître un jour donné de l'année. Avec ces probabilités, il sera maintenant possible de vérifier, à l'aide d'estimations basées sur des données réelles, l'impact d'assumer que tous les jours sont équiprobables sur le résultat final. Les données reçues couvrent les années 2020 à 2022 inclusivement. Étant donné qu'il n'y a que l'année 2020 qui est une année bissextile, les estimations seront effectuées en considérant qu'une année comporte 365 jours, le 29 février 2020 a donc été retiré des données. Ces estimations se rapprocheront tout de même de la réalité par rapport aux calculs précédents ; un calcul de probabilité comporte toujours certains raccourcis.

Pour arriver à son estimation, l'enseignante établit que la façon la plus efficace de procéder est de créer un petit programme informatique dont le résumé est présenté dans les prochaines lignes.

À partir de la banque de données des naissances, il est possible de calculer la probabilité de naître un jour donné, et ce, pour chaque jour de l'année. Cela donne des probabilités $p_1, p_2, \ldots, p_{365}$; la somme de toutes ces probabilités donne 1. Donc, si on choisit un groupe de personnes au hasard, la probabilité que le vecteur de leurs dates d'anniversaire soit $j = (j_1, \ldots, j_{30})$ est égale à

$$q_j = p_{j_1} \times \cdots \times p_{j_{30}}$$

Dénotons par S l'ensemble de tous les vecteurs j possibles ; cet ensemble est de cardinalité 360^{30}. Dénotons par C l'ensemble de tous les vecteurs comportant au moins deux valeurs j identiques. La probabilité qu'au moins deux personnes partagent leur date d'anniversaire est égale à

$$\sum_{j \in C} q_j \Big/ \sum_{j \in S} q_j$$

Étant donné que le dénominateur de cette fraction est 1, il ne faudrait, en théorie, que calculer la valeur du numérateur pour compléter le calcul ; malheureusement cela est difficile à faire en pratique. L'alternative est donc de générer aléatoirement un sous-ensemble de vecteurs T tel que $T \subset S$ et que T soit représentatif de la situation, en d'autres mots qu'il soit bien choisi. Pour ce faire, les vecteurs j sont créés en tenant compte des fréquences relatives des naissances quotidiennes ; cela suit l'idée *d'une simulation de Monte Carlo.* Comme T est un sous-ensemble de S, ses éléments sont distincts. Lorsqu'un nouveau vecteur est créé, s'il appartient déjà au sous-ensemble T, on l'ignore et on en génère un nouveau. Dans la situation, le nombre de vecteurs j créés fut de 10 000 ; cette valeur fut sélectionnée pour s'assurer que le programme s'exécute en un temps raisonnable tout en permettant d'effectuer l'expérience un grand nombre de fois.

Pour estimer la valeur de la probabilité, toujours selon la méthode de simulation de Monte-Carlo, on calcule alors le rapport,

$$\sum_{j \in (C \cap T)} 1 \Big/ \sum_{j \in (S \cap T)} 1 = |(C \cap T)| / |T| .$$

Est-ce que votre anniversaire est populaire ?

Avoir accès à une banque de données par jour sur plusieurs années nous permet de connaitre les dates d'anniversaire les plus communes et les moins communes. La banque de données que nous avons utilisée en est une du Québec couvrant les années 2020, 2021 et 2022. Il ne serait pas avisé de tirer des conclusions d'une base de données composées d'un aussi petit échantillon. Cependant, nous pouvons tout de même relever quelques tendances !

Rang	Moins fréquentes	Plus fréquentes
1	26 décembre	9 septembre
2	25 décembre	2 septembre
3	2 janvier	26 août
4	27 décembre	22 septembre
5	1 janvier	28 septembre

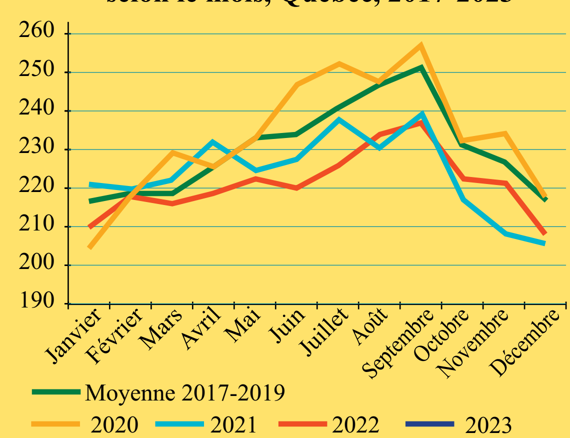
Le 29 février est la date la moins commune, mais ce tableau ne tient pas compte des années bissextiles.

Les dates du tableau s'expliquent par différents facteurs, dont le fait que certaines naissances ont lieu à des moments choisis par les hôpitaux et/ou les familles. Cela arrive lorsque celles-ci sont provoquées par exemple. Ainsi, on ne choisit pas de provoquer un accouchement dans les alentours d'un jour férié ; l'ensemble des dates les moins fréquentes sont près du jour de Noël ou du jour de l'An.

Les dates de naissance les plus populaires peuvent s'expliquer par le contexte neuf mois auparavant. Nous ne pouvons exclure que les naissances de la fin août jusqu'à la mi-septembre aient un lien avec la période des Fêtes neuf mois auparavant.

De façon plus générale, la météo semble avoir un impact étant donné que les mois d'hiver sont associés à des pics de naissances neuf mois plus tard, soit entre la fin de l'été et le début de l'automne. Cette tendance peut être observée dans le graphique ci-bas où l'évolution du nombre de naissances par mois est présentée. Il s'agit cette fois de données québécoises prises entre les années 2017 et 2023 pour lesquelles il nous faut remercier l'Institut de la statistique du Québec.

Nombre moyen de naissances par jour selon le mois, Québec, 2017-2023



Note : Données mensuelles détaillées disponibles sur le site Web de l'ISQ
Source : Institut de la statistique du Québec

Pour le programme C++, cela se résume donc à dénombrer le nombre de vecteurs qui sont composés d'au moins deux données identiques et à diviser ce nombre par 10 000 soit le nombre total de vecteurs créés.

L'enseignante, bien satisfaite de son programme, l'exécute et obtient que la probabilité qu'au moins deux élèves dans un groupe de 30 partagent leur anniversaire est d'environ 70,88 %.

À ce stade, l'enseignante ne peut que rire. Tout ce temps pour arriver à un résultat si près du premier qui était, rappelons-nous, d'environ 70,63 %. Quelques conclusions peuvent être tirées de cette petite aventure :

- Nous pouvons nous permettre de supposer que les dates d'anniversaire sont équiprobables dans les situations où la précision n'est pas de mise ;
- Nous pouvons apprécier davantage le paradoxe des anniversaires ;
- Un calcul de probabilité suppose souvent plusieurs raccourcis. Il faut savoir cibler ce qui nous intéresse et accepter une part d'imprécision.

Bien satisfaite de ses estimations, l'enseignante profita de la fin d'un cours pour présenter aux élèves son programme informatique ainsi que les résultats obtenus. Un élève se réjouit : « L'an prochain, dans mon nouveau groupe, j'ai plus de 70 % des chances d'avoir un ou une camarade de classe qui a le même anniversaire que moi puisque les classes ont 30 élèves ! »

Malheureusement, cette affirmation est fausse ! Il est vrai que selon le paradoxe des anniversaires, la probabilité qu'au moins deux élèves du groupe partagent leur anniversaire est légèrement supérieure à 70 %. Quelle est alors l'erreur dans le raisonnement de l'élève ? La probabilité s'amenuise grandement si on s'intéresse à la probabilité qu'une personne précise fasse partie des élèves qui partagent leur date d'anniversaire. Calculons cette probabilité. On commence par calculer la probabilité que tous les camarades de l'élève aient une date d'anniversaire différente de la sienne. On a vu qu'il y a 360^{30} combinaisons possibles pour les 30 dates d'anniversaire des élèves de la classe. Il a 365 choix pour la date d'anniversaire de l'élève et, pour les autres, il ne reste que 364 dates d'anniversaires possibles. Donc, au total, il y a 365 combinaisons. La probabilité que tous les futurs camarades de l'élève aient une date d'anniversaire différente de la sienne est donc

$$\frac{365 \times 364^{29}}{365^{30}} = \frac{364^{29}}{365^{29}} \approx 0{,}923\,5\ldots$$

La probabilité qu'un camarade partage sa date d'anniversaire est donc seulement de

$$\approx 1 - 0{,}9235\ldots \approx 7{,}65\%.$$

Le mot de la fin est laissé à l'enseignante :

« Soyez curieux, intéressez-vous aux probabilités et surtout gare aux conclusions rapides ! »

Simulation de Monte-Carlo

Pour s'assurer de la représentativité du sous-ensemble T, il faut que les vecteurs $j \in T$ soient créés en tenant compte des fréquences relatives des naissances quotidiennes. Une simulation de Monte-Carlo s'appuie sur la loi des grands nombres ; si on répète une expérience aléatoire un très grand nombre de fois et qu'on en fait la moyenne, ce qu'on obtient se rapproche de la réalité. Pour générer les vecteurs de façon aléatoire, on fractionne l'intervalle [0,1] en 365 sections en tenant compte des fréquences relatives des naissances quotidiennes ; chaque section représente un jour de l'année. Chaque intervalle sera défini ainsi (d_i, d_{j+1}) pour $i \in [0, 365]$ avec $d_1 = p_1$, $d_2 = p_1 + p_2$ et ainsi de suite jusqu'à $d_{365} = p_1 + p_2 + p_3 + \cdots + p_{365} = 1$.

Pour sélectionner une composante d'un vecteur $j \in T$, on pige au hasard un nombre $u \in [0,1]$ selon la loi uniforme. Si $u \in]d_{i-1}, d_i]$, on pose que la composante de u est i où i représente le i^e jour de l'année. Ce jour est ajouté au vecteur j des dates d'anniversaires. On répète la pige aléatoire 30 fois pour créer un premier vecteur j. En suivant ces mêmes étapes, il est alors possible de générer un très grand nombre de vecteurs $j \in T$.

Cette façon de procéder nous assure que le sous-ensemble $T \subset S$ est représentatif. De plus en prenant $|T| = 10\,000$, l'estimation sera précise à près de 95 % ce qui convient parfaitement !

Comment reproduire l'expérience aléatoire ?

Le programme informatique a été rédigé en C++ dans l'environnement *CLion*, mais le choix du langage de programmation et de l'environnement de développement est laissé à la discrétion de l'utilisateur. Pour reproduire le programme mentionné dans cet article, il faudra être en mesure ...

- d'importer efficacement des données ;
- de générer au hasard un nombre $u \in [0,1]$ selon la loi uniforme ;
- de parcourir une liste de nombres pour vérifier si celle-ci est composée de valeurs uniques ;
- de valider que toutes les listes sont différentes l'une de l'autre.

Les langages de programmation possèdent des modules intégrés au sein même de leur bibliothèque. Ces modules contiennent des fonctions, qu'on nomme ***méthodes*** en informatique, qu'il est possible d'utiliser. Des connaissances en programmation sont utiles pour recréer l'expérience de cet article, entre autres pour sélectionner les modules à importer pour avoir accès aux méthodes qui faciliteront l'écriture du programme.

Le paradoxe des anniversaires en mathématiques

Un paradoxe mathématique réfère à la situation où un raisonnement mathématique cohérent mène à un résultat qui semble illogique, contre-intuitif. Celui concernant les anniversaires est très connu en mathématiques ; voici comment on le présente normalement :

Combien de personnes doit-il y avoir dans un groupe pour que la probabilité qu'au moins deux personnes aient la même date d'anniversaire soit supérieure à 50 % ?

Contre-intuitivement, il ne faut que 23 personnes dans un groupe pour que cette probabilité soit supérieure à 50 %. Étant donné qu'une année possède 365 jours, en excluant les années bissextiles, il est naturel de penser qu'il faudrait un groupe de personnes beaucoup plus grand afin que cette probabilité atteigne le seuil de 50 %.

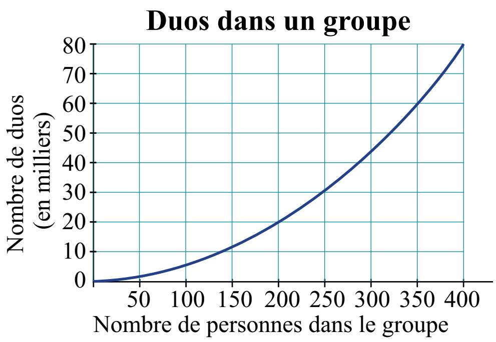
Intéressons-nous à toutes les possibilités de duos qu'il est possible de former dans un groupe de personnes. Si tous les élèves d'une même classe ont des dates d'anniversaire différentes, cela signifie que tous les duos qu'il est possible de former au sein de ce groupe seront composés d'élèves ayant des dates d'anniversaire distinctes. Pour la suite des explications, nous supposerons que les duos sont indépendants l'un de l'autre. Dans la réalité, le fait que les membres d'un duo aient des anniversaires distincts ou pas affecte les probabilités que d'autres élèves du même groupe partagent leurs anniversaires. Étant donné que la précision n'est pas de mise, mais que c'est plutôt le comportement de la situation qui nous intéresse, nous pouvons prendre ce raccourci. Nous supposerons également que les jours sont *équiprobables*. Nous pouvons calculer la probabilité que deux membres d'un même duo aient une date d'anniversaire distincte de la façon suivante :
$$\frac{365 \times 364}{365^2} = \frac{364}{365}.$$

En acceptant de considérer les duos indépendants pour la suite des calculs, la probabilité qu'au moins un duo soit composé de personnes partageant leur date d'anniversaire est

$$1-\left(\frac{364}{365}\right)^n, \text{ où } n \text{ est le nombre de duos.}$$

Plus le nombre de duos est élevé, plus cette probabilité se rapproche de 1. Le nombre de duos dans un groupe se calcule assez bien. Par exemple, dans un groupe de 30 personnes, chacune des 30 personnes peut être groupée avec l'une des 29 personnes restantes. Il y a donc 30×29 = 870 duos possibles. Cependant, il faut diviser ce résultat par 2, car il faut considérer que le duo composé de la personne A et B est le même que celui composé de la personne B et A. Il y a donc réellement 435 duos possibles dans un groupe de 30 personnes. Pourquoi le résultat est tout de même contre-intuitif alors qu'il est pourtant assez simple de calculer le nombre de duos ? En bref, notre cerveau nous joue des tours ! On n'anticipe pas ou on ne visualise pas bien une croissance de l'ordre n^2. Voici la courbe de croissance du nombre de duos pour un groupe de n personnes :

$$\text{Nombre de duos} = \frac{n(n-1)}{2}.$$

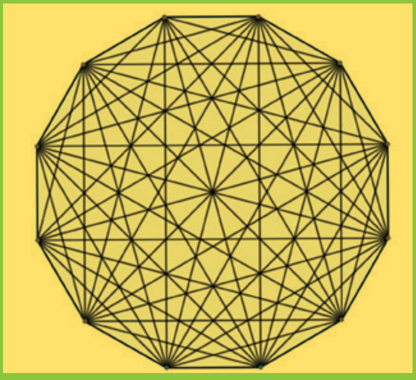


L'art et les duos

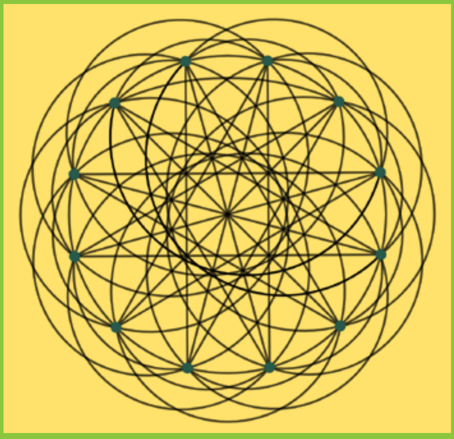
Existe-t-il une façon artistique de présenter tous les duos d'un groupe ? Supposons un groupe de 12 personnes et visualisons les duos au sein de ce groupe à l'aide d'un dodécagone régulier où chaque sommet représente une personne du groupe. Chaque segment reliant les sommets indique que deux personnes forment un duo.

$$\begin{aligned}\text{Nombre de duos} &= \frac{n(n-1)}{2}\\ &= \frac{12 \times 11}{2}\\ &= 66 \text{ duos.}\end{aligned}$$

Visuellement, dans le dodécagone régulier, cela représente tous ses côtés (12 côtés) ainsi que toutes ses diagonales (54 diagonales). Voici donc une représentation de tous les duos qu'il est possible de faire dans un groupe de 12 personnes :



Maintenant, laissons libre cours à notre imagination en conservant l'idée que chaque sommet, en bleu dans la figure ci-bas, représente une personne d'un groupe et qu'un segment reliant deux sommets implique que ces deux personnes forment un duo. Voici donc une autre façon de représenter l'ensemble des possibilités de duos dans un groupe de 12 personnes :



Il serait possible de répéter l'expérience à l'infini ! La créativité est sans limites !