ministère
éducation
nationale
ÉduSCOL
[dgesco]

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Mathématiques

Lycée

Ressources pour la classe de seconde

- Probabilités et Statistiques -

Juin 2009

Table des matières

I. Introduction

L'enseignement de la statistique et des probabilités constitue un enjeu essentiel pour la formation du citoyen, lui donnant des outils pour comprendre l'information chiffrée, décider et choisir de façon éclairée et participer au débat public. Il est par ailleurs très utile aux autres disciplines qui s'appuient fréquemment sur des modèles statistiques ou probabilistes. Le rapport de la commission de réflexion sur l'enseignement des mathématiques exprimait en ces termes l'enjeu de cet enseignement : « *l'objectif d'une initiation aux probabilités et à la statistique au niveau 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 programmes du collège ont inscrit l'étude des séries statistiques avec des indicateurs de position et de dispersion. Des notions de probabilité sont abordées en classe de troisième à partir de situations familières permettant, entre autres, de rencontrer des probabilités qui ne soient pas uniquement définies à partir de considérations intuitives de symétrie mais qui prennent appui sur l'observation d'épreuves répétées et la stabilisation des fréquences.

	Classe de sixième	Classe de cinquième	Classe de quatrième	Classe de troisième
Organisation et gestion de données	Organiser des données en choisissant un mode de représentation adapté. Lire et compléter une graduation sur une demi-droite graduée. Lire et interpréter des informations à partir d'une représentation graphique.	Repérage sur une droite graduée et dans le plan. Classes, effectifs, fréquences. Tableaux de données : lectures, interprétation, élaboration, représentations graphiques.	Moyenne pondérée.	Caractéristiques de position : médiane, quartiles. Approche de caractéristiques de dispersion : étendue. Notion de probabilité

En conséquence, il s'agit de veiller à bien inscrire l'enseignement de la classe de seconde en continuité avec celui du collège. Cela vaut autant pour la seconde générale et technologique que pour la voie professionnelle dont les programmes en vigueur à la rentrée 2009 font une part importante aux probabilités et à la statistique[1]. Il conviendra notamment d'éviter des révisions systématiques et de proposer des situations permettant le réinvestissement des notions abordées dans les classes précédentes.

Le travail statistique sur données réelles, brutes ou préalablement traitées avec l'aide incontournable de l'outil informatique, est de nature à favoriser la prise d'initiative et la conduite de raisonnements pour interpréter, analyser ou comparer des séries statistiques sur des sujets en prise avec l'actualité des élèves. Pour reprendre le rapport de la commission de réflexion sur l'enseignement des mathématiques, « *le matériau brut travaillé par la statistique est constitué de données expérimentales, les outils théoriques utilisés sont essentiellement la géométrie et l'algèbre linéaire pour la statistique exploratoire et les probabilités pour la statistique inférentielle et l'outil matériel est l'ordinateur* ».

Ainsi les outils de statistique descriptive travaillés au collège pourront être efficacement sollicités lors de l'exploration de fichiers comportant des données réelles afin d'exhiber des

[1] BO spécial n°2 du 19/02/2009

régularités, des dominantes ou des caractéristiques que la masse de données ou le nombre de variables étudiées ne livrent pas facilement. Cette pratique de fouille de données (statistique exploratoire, data mining) est de nos jours essentielle tant les données disponibles sont nombreuses et constituent des repères indispensables pour les entreprises ou les politiques publiques dans leurs choix stratégiques et décisionnels. Histogrammes, représentations graphiques, calculs de moyennes, de médianes, de quartiles sont autant d'outils que les élèves pourront utiliser pour faire émerger des questions sur les données, dégager des problématiques d'étude ou résumer l'information essentielle, d'autant plus que les outils informatiques accessibles aux élèves permettent de travailler sur des gros fichiers.

Les premiers éléments de probabilité ont été abordés au collège essentiellement dans des situations de jeux (lancers de dés ou de pièces, loteries, tirages dans des urnes). Cela a permis une première approche de quelques lois de probabilité qui seront progressivement décontextualisées au lycée en vue de fournir des modèles pour d'autres champs d'application, tant dans les domaines scientifiques (sciences physiques et chimiques, sciences de la vie et de la terre, sciences de l'ingénieur, etc.) que dans les sciences économiques et sociales. Ainsi les élèves ont été familiarisés, par des lancers de pièces équilibrées, avec la loi uniforme sur {0 ; 1} qui permettra, par exemple, d'aborder de nombreux sujets (sex-ratio, parité). De même le tirage dans une urne composée de boules de deux couleurs différentes, dans des proportions p et $(1 - p)$, fournit une première approche de la loi de Bernoulli, qui s'ouvrira sur des applications aux sondages (estimations à la sortie des urnes les soirs d'élection), au contrôle de qualité des productions industrielles (maîtrise statistique des processus de production) ou aux diverses estimations sur échantillon. Toutes ces questions relèvent de la statistique inférentielle, qui fonde ses résultats sur des considérations probabilistes et permet l'induction à partir de données observées.

Les questions de fluctuation d'échantillonnage constituent un axe important de la formation du futur citoyen, qui aura ainsi été sensibilisé au lycée à la nécessaire prudence à avoir avant d'interpréter une évolution ou d'effectuer des comparaisons. En effet toute évolution de moyenne ou de proportion, toute comparaison d'échantillons doit être nuancée et relativisée au regard des variations liées à la fluctuation d'échantillonnage.

Afin d'entrer vraiment dans une démarche statistique en lien avec les concepts probabilistes, on gagnera à utiliser, comme fil rouge, un fichier de données réelles pour mettre en œuvre ou pour illustrer les différentes notions inscrites au programme[2]. En procédant ainsi, on limite le temps d'appropriation des données et les élèves peuvent plus rapidement se concentrer sur les outils mathématiques, la situation étudiée devenant familière.

Par ailleurs, c'est en ayant recours à des données réelles que l'on développe les capacités d'observation et de raisonnement des élèves : comprendre la nature des données, repérer l'organisation d'un tableau, imaginer et réaliser des représentations ou des calculs adaptés, comprendre un graphique sont autant d'occasions de raisonner et d'exercer l'esprit critique. Dans les exemples développés dans ce document, on a tenu à souligner ces différents aspects de la formation, montrant l'ampleur et l'ambition des raisonnements conduits, ainsi que la place légitime de cet enseignement dans le programme de mathématiques.

2 Voir le dossier proposant une progression et des activités utilisant le fichier des populations des communes françaises : http://www.ac-grenoble.fr/maths/docresseconde/Exemple%20de%20progression.zip (fichiers à laisser dans le même dossier)

II. Des statistiques aux probabilités

1. Statistique descriptive, analyse de données

1.1. Résumé des notions abordées au collège

Les notions de moyenne, médiane, étendue, quartiles et écart interquartile ont été développées au collège ainsi que leurs interprétations.

Ces notions pourront être sollicitées en classe de seconde dans plusieurs domaines en lien avec les autres disciplines, par exemple pour étudier des séries de mesures expérimentales en sciences physiques. Il est alors possible de traiter plusieurs questions autour de l'intervalle interquartile, de son amplitude, etc.

1.2. Analyse de données

La classe de seconde est l'occasion d'une part de consolider l'utilisation des fonctions statistiques des calculatrices et d'autre part de traiter, à l'aide d'un tableur, des séries statistiques riches et variées comportant un grand nombre de données brutes en lien avec des situations réelles. À titre d'exemples, on pourra trouver de telles données :

- sur le site de l'INSEE , population des régions, départements et communes :
http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/populations-legales/france-departements.asp

- sur le site de l'INED, espérance de vie à la naissance :
http://www.ined.fr/fichier/t_telechargement/18154/telechargement_fichier_fr_sd2006_t2_fm.xls

- sur le site de météo France ou des sites de particuliers tel :

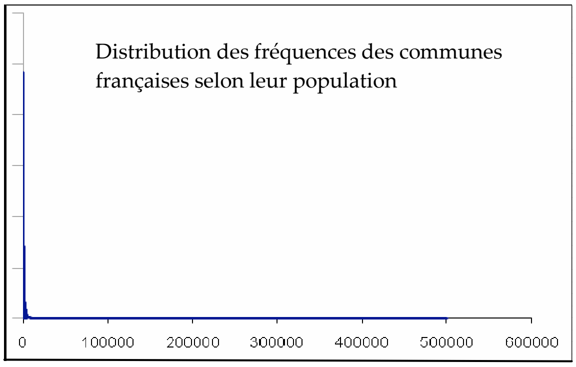
http://www.meteociel.com/climatologie/climato.php

Par ailleurs on pourra exploiter, dans le cadre de travaux interdisciplinaires, des données issues d'autres disciplines (sciences physiques, sciences de la vie et de la terre, sciences de l'ingénieur etc.).

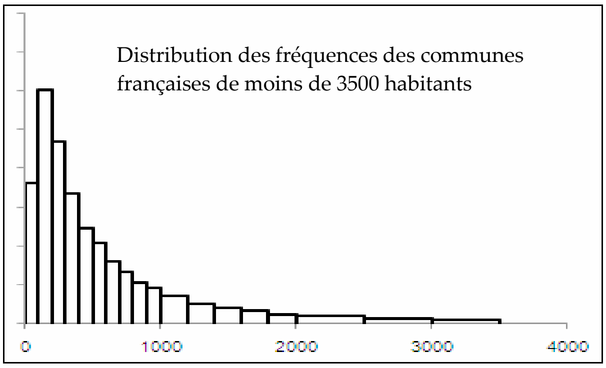
Exemple : fichier des communes françaises
Le nombre important de données peut inciter, dans un premier temps, à effectuer une représentation graphique (graphique 1).

Le résultat obtenu est surprenant et soulève immédiatement la question de la répartition des communes en fonction de leur nombre d'habitants, suggérant de scinder la série, par exemple en isolant les communes de moins de 3500 habitants qui est le seuil retenu dans la loi électorale de 2007. Le graphique 2 donne la répartition de ces communes[3].

[3] Réalisation d'histogrammes à classes d'amplitudes inégales, on pourra consulter : http://www.ac-grenoble.fr/maths/guppy/pages/fiches/Mediane/Tableur_stats_1_Var.htm



Graphique 1

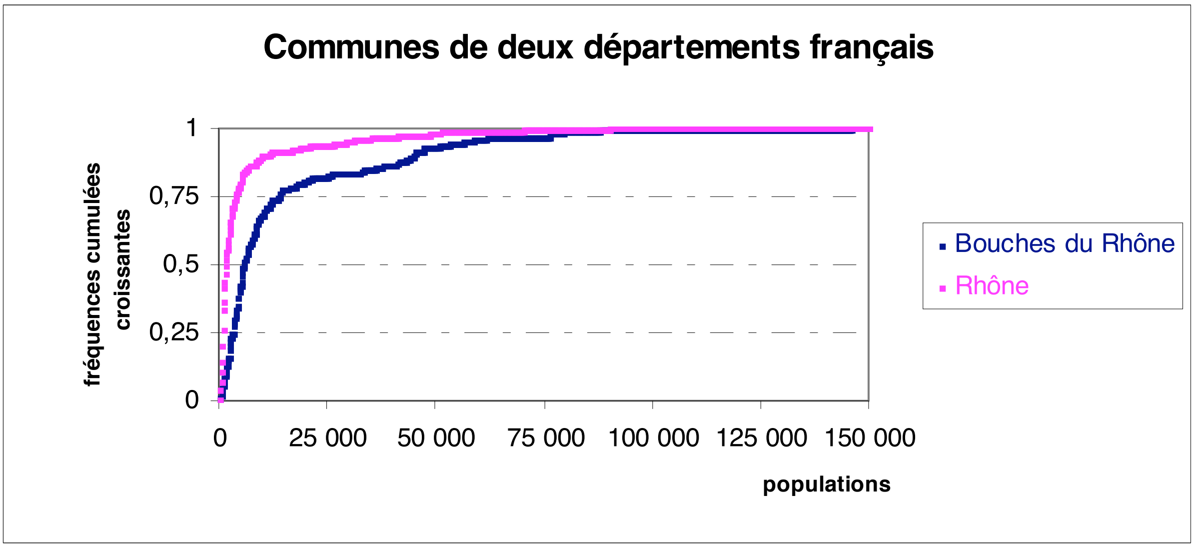


Graphique 2

Il est clair, en observant le graphique 1 ci-dessus que la moyenne (1760 habitants) n'est pas pertinente pour résumer cette série statistique. On lui préfèrera ici la médiane et les quartiles.

1.3. Fréquences cumulées croissantes

La courbe des fréquences cumulées croissantes[4] permet de représenter la distribution d'une série statistique ainsi que l'illustre l'exemple ci-dessous.
Exemple : comparaison entre deux départements[5]
Le graphique ci-après est réalisable à partir du fichier INSEE des régions, départements et communes de France. La compréhension des deux courbes conduit les élèves à des raisonnements formateurs pour répondre aux questions suivantes :
- comment interpréter l'antécédent[6] de 0,5 (resp 0,25 ; 0,75) par chacune de ces fonctions ?
- l'une des courbes est située en dessous de l'autre ; comment interpréter cette propriété ?



---

[4] Elle permet de retrouver une valeur approchée de la médiane, mais ce n'est pas une méthode à préconiser lorsque l'objectif est uniquement de déterminer une valeur médiane.
[5] Activité à l'adresse suivante :
http://www.ac-grenoble.fr/maths/docresseconde/Exemple de progression/activite_communes.doc
[6] Sous l'hypothèse que les points aient été reliés.

2. Probabilité sur un ensemble fini

Le travail sur les probabilités, initié en classe de troisième, est stabilisé et consolidé en classe de seconde avec, en perspective, une démarche de modélisation de phénomènes réels.

2.1. Résumé des notions abordées en troisième

Notion de probabilité, calcul dans des situations familières (lancer de pièces ou de dés, roue de loterie, urnes). Probabilités estimées par des fréquences observées sur de longues séries. Application pour modéliser des situations de la vie courante. Les expériences aléatoires concernent des situations à une ou deux épreuves qui n'excèdent pas 6 éventualités au total.

2.2. Distribution de probabilité sur un ensemble fini, probabilité d'un événement

Il s'agit dans un premier temps de consolider les notions abordées en classe de troisième.

Une distribution de probabilité sur un ensemble Ω est définie par la donnée des probabilités des éléments de Ω. Un événement est défini comme sous-ensemble de Ω. C'est cette définition ensembliste qui permet de calculer la probabilité d'un événement en ajoutant les probabilités des éléments qui le constituent. On consolide à l'occasion la notion d'ensemble ou de sous-ensemble, ce qui permet, entre autres, d'ancrer l'idée que dans un ensemble on ne répète pas les éléments et que leur ordre n'importe pas.

Les distributions de probabilité peuvent être estimées par observation de la stabilisation des fréquences[7] sur de longues séries d'expériences[8] ou bien par des considérations géométriques ou physiques en référence à l'équiprobabilité.

De même qu'au collège les élèves ont utilisé, sans formalisme, quelques éléments de langage sur les probabilités[9] , de même en classe de seconde certaines notations usuelles seront utilisées pour leur commodité et sans donner lieu à un formalisme excessif :
$\overline{A}$, $p(\{1,2,3\})$, $p(A \cap B)$ ou $p(A \text{ et } B)$, $p(A \cup B)$ ou $p(A \text{ ou } B)$, $\text{card}(A)$.

2.3. Modélisation, modélisations ?

La simulation a pour préalable de choisir un modèle.

Exemple 1 : somme de deux dés
Cette situation se prête volontiers à la mise en œuvre d'une démarche consistant à proposer un modèle et à le confronter aux données d'expérience. Les résultats compris entre 2 et 12 peuvent conduire certains élèves à faire porter l'équiprobabilité sur l'ensemble des 11 résultats observables. Par quelques expérimentations avec des dés, puis en ayant recours à des simulations, on est conduit à rejeter ce modèle pour proposer de faire porter l'équiprobabilité sur les 36 couples de résultats {(1,1) ;(1,2) …(6,6)}, présentés usuellement sous forme de tableau croisé.

[7] On en restera à la perception intuitive de la loi des grands nombres.
[8] On trouvera un exemple d'approche fréquentiste page 13 dans le document ressource des nouveaux programmes de lycée professionnel http://www.ac-grenoble.fr/maths/docresseconde/Proba_stat_LP.doc
[9] Cf. document ressource du collège :
http://www.ac-grenoble.fr/maths/docresseconde/doc_ressource_clg_probabilites.pdf

D'autres exemples[10] pourront être développés, montrant aux élèves les cheminements de pensée entre les modèles retenus et les expériences réelles, et posant la question de l'ensemble sur lequel porte l'équiprobabilité.

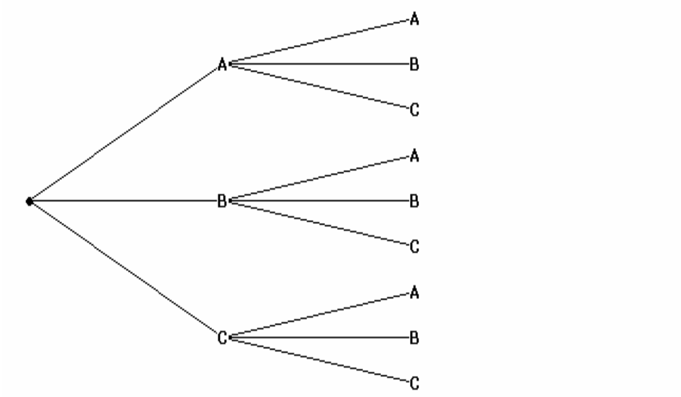
Exemple 2 : le problème des deux personnes qui s'assoient "au hasard "

Un square est équipé de trois bancs à deux places. Deux personnes arrivent successivement et s'installent au hasard. Quelle est la probabilité que ces personnes soient assises côte à côte ?

En faisant l'hypothèse de l'équiprobabilité des issues :

Première modélisation[11] : on numérote les six places 1, 2, 3, 4, 5 et 6, chaque paire représente les deux places occupées. En comptant les paires {1,2}, {1,3} … {5, 6}, on obtient pour cet événement une probabilité de $\frac{3}{15}$ soit $\frac{1}{5}$.

Deuxième modélisation[12] *:* on note les trois bancs A, B et C, les résultats de l'expérience peuvent être codés par des couples, par exemple (B,A) : la première personne s'assoit sur le banc B et la deuxième sur le banc A. On peut aussi construire l'arbre des possibles :



On obtient une probabilité de $\frac{1}{3}$.

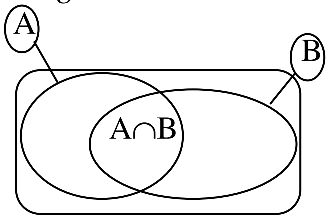
Ces deux modèles donnent des résultats différents, « tirer au hasard » n'étant pas une information suffisante, même en rajoutant une hypothèse d'équiprobabilité. Une description plus précise de l'expérience doit être fournie pour induire un choix de modèle.

3. Calculs de probabilités

3.1. Réunion et intersection de deux événements

Les symboles d'union et intersection sont introduits en liaison avec les conjonctions « ou » et « et », en comparant leur sens mathématique avec leur usage dans la langue commune.

La formule $p(A \cup B) + p(A \cap B) = p(A) + p(B)$ pourra être illustrée par un diagramme de Venn :



[10] Le problème du Duc de Toscane : lancer de trois dés, quelle est la somme la plus probable ? : http://www.ac-grenoble.fr/maths/docresseconde/Le paradoxe du Duc de Toscane.doc

[11] On peut imaginer que le choix de la place s'effectue par tirage sans remise dans une urne avec 6 jetons numérotés de 1 à 6.

[12] Le choix du banc pourrait s'effectuer par tirage avec remise dans une urne contenant 3 jetons A, B et C.

Exemple : avec le fichier des communes françaises
On choisit au hasard une commune dans l'ensemble des communes de France.
Quelle est la probabilité de l'événement A : « cette commune est dans votre région » ?
Quelle est la probabilité de l'événement B : « sa population est inférieure à 1000 habitants » ?
Définir les événements $A \cup B$ et $A \cap B$ puis calculer leur probabilité.
Cet exemple, à partir d'un fichier comportant de très nombreuses données, peut conduire les élèves à pratiquer des instructions logiques dans les conditions appelées par la fonction NB.SI.

3.2. Tableaux croisés

Les tableaux croisés rencontrés dans des résumés d'enquêtes se prêtent à des calculs simples de probabilités d'intersections ou d'union[13].

3.3. Arbres des possibles

Les arbres décrivant de façon exhaustive les issues d'une expérience ont pu être abordés en classe de troisième ; on peut consolider cette pratique pour aider les élèves à se construire des images mentales fiables et être plus assurés dans les modélisations et les calculs.

Ces arbres aident au dénombrement et sont des supports de raisonnement. Il n'est pas toujours nécessaire ni matériellement possible d'en représenter toutes les branches. On peut développer les capacités d'abstraction des élèves en utilisant des pointillés dans leur construction.

Exemple : probabilité d'avoir la même date anniversaire[14]
En regardant les dates anniversaires des élèves dans les classes du lycée, on peut être surpris du nombre de classes dans lesquelles deux élèves fêtent leur anniversaire le même jour.

Ce résultat étonnant peut inciter à faire un calcul de probabilité.

On pourra proposer de commencer par un exemple plus simple :

« Dans un groupe de quatre personnes prises au hasard, quelle est la probabilité qu'au moins deux d'entre elles fêtent leur anniversaire le même mois ? On suppose que, pour chaque personne, tous les mois d'anniversaire sont équiprobables et on les numérote de 1 à 12. ».

On peut reformuler ce problème en assimilant l'expérience à un tirage aléatoire dans une urne : « une urne contient 12 boules numérotées de 1 à 12, on effectue au hasard et avec remise quatre tirages successifs, et l'on note les numéros obtenus, dans l'ordre d'apparition. Chaque numéro tiré correspond au mois d'anniversaire d'une des personnes ».

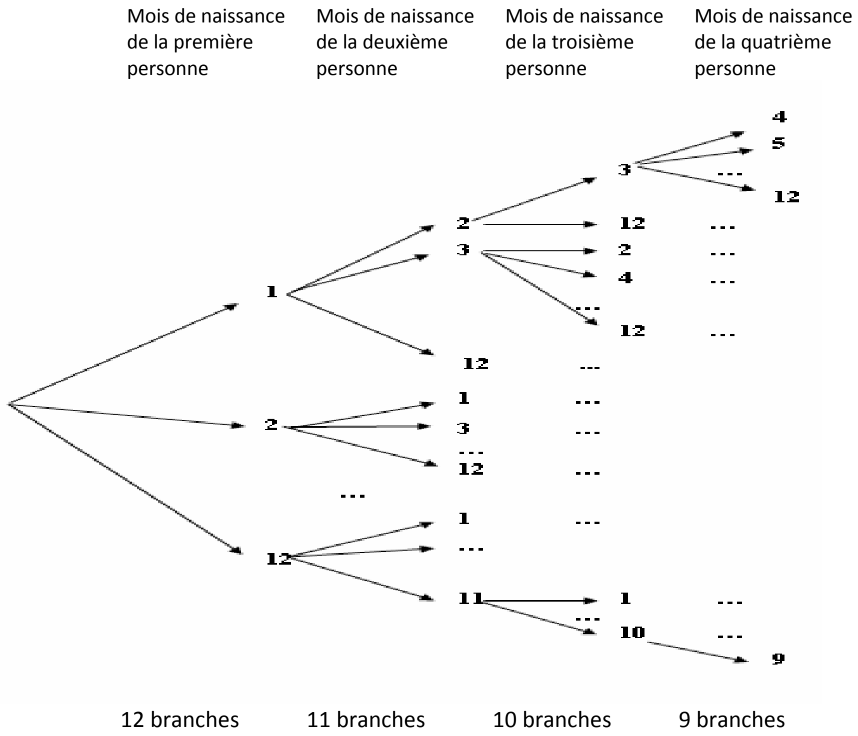
On peut compter le nombre total des issues avec un arbre comportant des pointillés.

Ensuite on peut rechercher le nombre d'éléments de l'événement étudié, montrer la difficulté que l'on rencontre pour décrire et compter directement les cas favorables, puis faire réfléchir à l'intérêt et à l'énoncé de l'événement contraire (négation de "au moins…").

---

[13] On peut utiliser des tableaux croisés disponibles sur le site de l'INSEE, par exemple le diplôme le plus élevé selon l'âge et le sexe : http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=NATCCF07235
[14] Cf. feuille de calcul
http://www.ac-grenoble.fr/maths/docresseconde/Anniversaires_probabilit%E9s_simulation.xls

Illustration de l'arbre des possibles de l'événement contraire:



Pour bien s'assurer de la compréhension de l'arbre, on peut suivre un trajet de la racine à une extrémité de branche et demander aux élèves d'interpréter le résultat par une phrase.

On trouve au total $12\times 11\times 10\times 9 = 11880$ issues possibles. Ce qui donne comme probabilité de l'événement contraire : $\dfrac{12\times 11\times 10\times 9}{12^4}$, soit environ 0,43 ; d'où la probabilité cherchée (environ 0,57). On peut ensuite adapter ces calculs de probabilité pour un groupe de cinq, puis six personnes[15].

Retour au problème initial : le même raisonnement, immédiatement transposé, permet de résoudre la situation des mêmes dates anniversaires et de rechercher la taille du groupe de personnes pour avoir une probabilité supérieure à 0,8 qu'au moins deux d'entre elles fêtent leur anniversaire à la même date.

Utilisation d'un algorithme

Si n désigne le nombre de personnes du groupe, il s'agit de déterminer à partir de quelle valeur de n le nombre $q=\dfrac{365\times 364\times\cdots\times(365-(n-1))}{365^n}$ est inférieur à 0,2, avec $q=1-p$.

En remarquant que q peut s'écrire comme une répétition de multiplications : $q=\dfrac{365-0}{365}\times\dfrac{365-1}{365}\times\dfrac{365-2}{365}\times\cdots\times\dfrac{365-(n-1)}{365}$, on peut élaborer un algorithme de calcul de ce nombre selon la valeur de n (algorithme 1), puis par essais successifs, déterminer la première valeur de n qui répond à la question[16].

---

[15] Réponses: 0,62 et 0,78.

[16] Pour $n = 35$, on trouve $p\approx 0{,}814$, alors que pour $n = 34$ on a $p\approx 0{,}795$.