Le paradoxe des dates d'anniversaires

Guillaume Blanc
LAL

4 juillet 2017

1 Ce qui semble « extraordinaire » peut être assez probable

Notre esprit est peu à même d'estimer la probabilité d'occurences d'événements. Il est le siège d'un certain nombre de biais dits « cognitifs », des erreurs de raisonnement, qui faussent notre jugement. Notre raisonnement intuitif (le système 1 selon Daniel Kahneman, voir [2]) fait un certain nombre de raccourcis pour interpréter le monde en permanence, des « heuristiques », opérations mentales automatiques qui fonctionnent dans la quasi-totalité des cas, mais qui nous amènent à tomber dans un certain nombre de pièges dans certains cas, quand le cerveau (système 2) du raisonnement réfléchi ne prend pas le relais suffisamment rapidement. L'estimation des probabilités fait parti de ces écarts. Ainsi des événements peuvent nous paraître très improbables alors qu'en réalité ils sont suffisamment probable pour avoir lieu. Cela donne lieu à des interprétations fantaisistes faisant appel à l'extraordinaire, alors qu'en fait, un (plus ou moins) simple calcul de probabilité permet de faire la lumière. Ainsi en est-il du paradoxe des anniversaires.

2 Coïncidence des jours d'anniversaires dans un groupe

Ne nous paraît-il ainsi pas très improbable que dans un groupe de taille raisonnable, comme un groupe de TD à l'université composé typiquement d'une vingtaine d'étudiants, on trouve deux personnes dont l'anniversaire tombe le même jour ? Ce qui suit est inspiré de [1].

3 Problème

Calculons donc la probabilité pour que deux personnes au moins, dans un groupe de N personnes, aient leur anniversaire le même jour.

Pour cela, il faudrait calculer la probabilité que 2 personnes seulement parmi les N du groupe aient leur anniversaire le même jour (voir paragraphe 6), que l'on ajoute à la probabilité que 3 personnes seulement aient leur anniversaire le même jour, que l'on ajoute à la même probabilité que 4 personnes, sans oublier les couples qui peuvent avoir leur anniversaire le même jour, les triplets, les quadruplets, etc.

Il est plus simple de considérer le cas complémentaire : la probabilité qu'au moins 2 personnes dans un groupe de N personnes aient leur anniversaire le même jour plus la probabilité qu'aucune personne dans un groupe de N n'ait son anniversaire le même jour doit être égal à un : on considère toutes les occurrences possibles, la probabilité résultante est donc de un, certaine.

On peut ainsi, pour résoudre notre problème, calculer plutôt cette probabilité complémentaire, à savoir celle qu'aucune personne dans un groupe de N n'ait son anniversaire le même jour.

4 Démonstration

Considérons la première personne du groupe, elle a 365 possibilités pour son anniversaire. Comme on veut qu'aucune personne du groupe n'ait la même date, la deuxième personne a donc 364 possibilités. La N^{e} personne n'a donc plus que $(365 - N + 1)$ possibilités. Le nombre de possibilités pour l'ensemble du groupe est donc :

$$365 \times 364 \times 363 \times ... \times (365 - N + 1)$$

Comme l'ordre dans lequel chaque personne « choisit » sa date anniversaire importe peu, il faut diviser ce nombre de possibilités par le nombre total de possibilités, et pas seulement celles concernant les dates différentes. Comme au total chaque personne a 365 choix, pour N personnes, cela fait donc 365^N possibilités.

Donc la probabilité que personne n'ait son anniversaire le même jour est :

$$\overline{P_{\geq 2}(N)} = \frac{365 \times 364 \times 363 \times ... \times (365 - N + 1)}{365 \times 365 \times 365 \times ... \times 365}$$

soit :

$$\overline{P_{\geq 2}(N)} = \frac{365!}{(365 - N)!} \times \frac{1}{365^N}$$

Donc la probabilité cherchée, à savoir celle que deux personnes *au moins* aient leur anniversaire le même jour est :

$$P_{\geq 2}(N) = 1 - \overline{P_{\geq 2}(N)} = 1 - \frac{365!}{(365 - N)!} \times \frac{1}{365^N} \tag{1}$$

5 Application

On peut écrire un bout de programme en python pour calculer ces probabilités :

```
import numpy as np
import pylab as pl

nb = 100 # nombre de personnes dans le groupe
nb_jour = 365 # nombre de jours dans l'annee

tab_n   = np.linspace(0,nb,nb, dtype=int)
tab_p0  = np.zeros(nb)

# probabilite que personne parmi N n'ait son anniversaire le meme jour
tab_p0[0] = 1
for i in range(1,nb):
    tab_p0[i] = tab_p0[i-1] * float(nb_jour-i+1)/float(nb_jour)

pl.plot(tab_n, 1-tab_p0, 'k-')

pl.show()
```

L'idée étant de calculer numériquement cette probabilité, non pas en passant par les factorielles, qui donnent des rapports de nombres trop grands pour être facilement calculés, mais en incrémentant les rapports de probabilités un à un.

Le résultat est donné sur la figure 1. Le tableau 2 donne quelques valeurs de la probabilité pour quelques tailles du groupe de personnes. On se rend compte que ces probabilités sont élevées. En particulier, elles dépassent 0.5 à partir de seulement 23 personnes ! Il n'est donc pas rare, en dépit de notre « intuition », que deux étudiants aient leur anniversaire le même jour dans un groupe de TD.

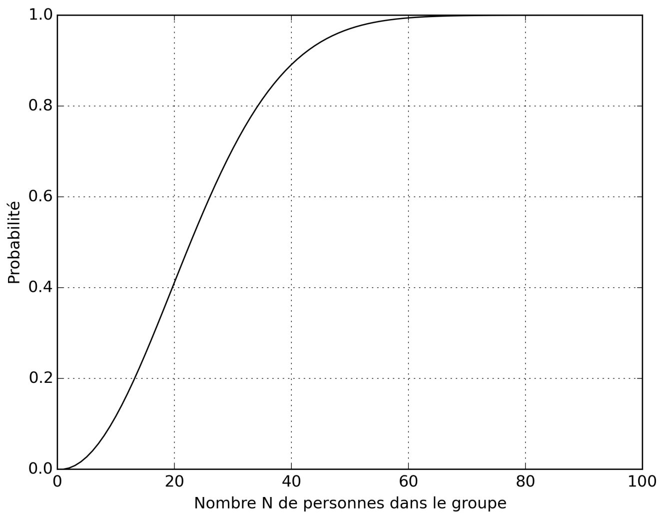

FIGURE 1 – Probabilité d'avoir au moins deux personnes ayant leur anniversaire le même jour en fonction de la taille du groupe.

N	$P_{\geq 2}(N)$
10	0.12
20	0.41
22	0.48
23	**0.51**
30	0.71
40	0.89
50	0.97
60	0.994
70	0.999

FIGURE 2 – Probabilité qu'au moins deux personnes dans un groupe de N personnes aient leur anniversaire le même jour. On remarque que cette probabilité est élevée, elle est de plus de 0.5 pour 23 personnes.

6 Pour aller plus loin...

On peut essayer de trouver la probabilité que *seulement* deux personnes dans un groupe de N personnes aient leur anniversaire le même jour, toutes les autres personnes ayant donc une date d'anniversaire différente [3].

6.1 Exemple

Pour trouver cette probabilité, il m'a fallu cogiter un peu. Afin de pourvoir représenter les différents cas en jeu, j'ai réduit le problème aux sept jours de la semaine. Problème qui devenait ainsi celui de savoir quelle est la probabilité que dans un groupe de N personnes, deux personnes et seulement deux aient leur anniversaire le même jour de la semaine.

Ce problème est exactement le même que celui de tirer N fois une boule numérotée (de 1 à 7 pour les jours de la semaine, ou bien de 1 à 365 pour les jours de l'année) avec remise, où l'on cherche à savoir quelle est la probabilité de tirer seulement deux fois la même boule.

Pour $N = 2$ personnes, la première a 7 possibilités, tandis que la deuxième n'en a qu'une. Tandis que l'éventail total des possibilités est de 7 pour la première personne et 7 pour la seconde, soit 7^2 en tout.

Donc :

$$P_2^7(2) = \frac{7 \times 1}{7 \times 7} = \frac{1}{7} \simeq 0.14$$

On peut vérifier que :

$$P_2^7(2) = 1 - P_{\geq 2}^7(2) = 1 - \frac{7 \times 6}{7 \times 7} = \frac{1}{7}$$

Pour $N = 3$ personnes, la première a 7 possibilités, la deuxième en a 1, la troisième en a 7-1 $=$ 6. Mais cette fois-ci il faut tenir compte de la répartition des « tirages » : si on raisonne avec les boules numérotées que l'on tire dans une urne, il y a différentes façons de tirer trois boules dont deux fois la même. On représente ci-dessous chaque tirage de 3 boules, les boules identiques en noir, et les autres en blanc :

$$\begin{array}{ccc} \bullet & \bullet & \circ \\ \bullet & \circ & \bullet \\ \circ & \bullet & \bullet \end{array}$$

Ce qui nous donne 3 possibilités. Ce qui revient à déterminer le nombre de combinaisons de 2 parmi un ensemble de 3 : $\binom{3}{2}$. Ainsi la probabilité cherchée est :

$$P_2^7(3) = \binom{3}{2} \frac{7 \times 1 \times 6}{7 \times 7 \times 7} = 3 \times \frac{6}{49} = \frac{18}{49} \simeq 0.367$$

On peut vérifier cela avec le fait que la probabilité qu'au moins deux personnes parmi 3 aient leur anniversaire le même jour de la semaine est aussi égal à la probabilité que seulement 2 personnes parmi 3 aient leur anniversaire le même jour de la semaine plus la probabilité que les trois personnes aient leur anniversaire le même jour de la semaine :

$$P_2^7(3) + P_3^7(3) = 1 - P_{\geq 2}^7(3)$$

Avec :

$$P_3^7(3) = \binom{3}{3} \frac{7 \times 1 \times 1}{7 \times 7 \times 7} = \frac{1}{49} \simeq= 0.02$$

et :

$$P_{\geq 2}^7(3) = \frac{7 \times 6 \times 5}{7 \times 7 \times 7} = \frac{30}{49} \simeq 0.61$$

Donc :

$$\frac{18}{49} + \frac{1}{49} = \frac{19}{49} = \frac{49 - 30}{49}$$

Regardons le cas de 4 personnes. Il faut s'intéresser aux combinaisons de 2 parmi 4, de 3 parmi quatre et de 4 parmi 4. Ces combinaisons sont représentées ci-dessous :

2 parmi 4 :

$$\begin{array}{cccc} \bullet & \bullet & \circ & \circ \\ \bullet & \circ & \bullet & \circ \\ \bullet & \circ & \circ & \bullet \\ \circ & \bullet & \bullet & \circ \\ \circ & \bullet & \circ & \bullet \\ \circ & \circ & \bullet & \bullet \end{array}$$

$$\binom{4}{2} = 6$$

3 parmi 4 :

$$\begin{array}{cccc} \bullet & \bullet & \bullet & \circ \\ \bullet & \bullet & \circ & \bullet \\ \bullet & \circ & \bullet & \bullet \\ \circ & \bullet & \bullet & \bullet \end{array}$$

$$\binom{4}{3} = 4$$

4 parmi 4 :

$$\begin{array}{cccc} \bullet & \bullet & \bullet & \bullet \end{array}$$

$$\binom{4}{4} = 1$$

Or :

$$P^7_{\geq 2}(4) = 1 - \frac{7 \times 6 \times 5 \times 4}{7 \times 7 \times 7 \times 7} = \frac{223}{343} \simeq 0.65$$

mais aussi :

$$\begin{aligned} P^7_2(4) + P^7_3(4) + P^7_4(4) &= \binom{4}{2}\frac{7 \times 1 \times 6 \times 5}{7 \times 7 \times 7 \times 7} + \binom{4}{3}\frac{7 \times 1 \times 1 \times 6}{7 \times 7 \times 7 \times 7} + \binom{4}{4}\frac{7 \times 1 \times 1 \times 1}{7 \times 7 \times 7 \times 7} \\ &= 6 \times \frac{30}{343} + 4 \times \frac{6}{343} + \frac{1}{343} = \frac{205}{343} \end{aligned}$$

Or cette dernière expression ne coïncide pas avec la précédente. Pour la faire coïncider (voir [3]) il faut ajouter la probabilité que deux couples de personnes aient leur anniversaires le même jour de la semaine (deux personnes un jour donné, et deux autres personnes un autre jour). Pour cela, il faut choisir deux jours différents parmi les 7 jours de la semaine, soit $\binom{7}{2}$, puis il faut choisir le premier couple, 2 choix parmi 4, et le deuxième couple, 2 choix parmi 2. La probabilité associée est donc :

$$P^7_{2\times 2}(4) = \binom{7}{2} \times \binom{4}{2} \times \binom{2}{2} \times \frac{1}{7^4} = \frac{21 \times 6 \times 1}{7^4} = \frac{18}{343}$$

Or :

$$\frac{18}{343} + \frac{205}{343} = \frac{223}{343}$$

On retrouve ainsi $P^7_{\geq 2}(4)$.

6.2 Revenons aux anniversaires

On a donc une formule générale pour la probabilité que deux personnes exactement aient leur anniversaire le même jour :

$$P^{365}_2(N) = \binom{N}{2} \times \frac{365 \times 1 \times 364 \times 363 \times ... \times (365 - N + 2)}{365 \times 1 \times 365 \times 365 \times ... \times 365} = \binom{N}{2} \times \frac{365!}{(365 - N + 1)!} \times \frac{1}{365^N}$$

Ou bien, de manière plus générale, pour J personnes parmi N ayant leur anniversaire le même jour :

$$P^{365}_J(N) = \binom{N}{J} \times \frac{365!}{(365 - N + J - 1)!} \times \frac{1}{365^N} \tag{2}$$

Mais si on veut décomposer la probabilité initiale « au moins deux personnes du groupe ont leur anniversaire le même jour, » il faut tenir compte non seulement des probabilités de J personnes exactement ayant le même anniversaire mais également des couples de personnes qui ont leur anniversaire le même jour, ainsi que des triplets de personnes, des quadruplets, etc.

6.2.1 Cas des couples

Il y a $N/2$ couples, au maximum susceptibles de partager un anniversaire.

On a vu dans l'exemple ci-dessus la probabilité que deux couples parmi 4 personnes aient chacun un anniversaire un jour de la semaine commun à chacun. On peut généraliser cela.

La probabilité que deux couples (parmi N personnes) exactement partagent chacun une date anniversaire est donc :

$$P^{365}_{2^2}(N) = \binom{365}{2} \times \binom{N}{2} \times \binom{N-2}{2} \times \frac{363 \times 362 \times ... \times [365 - (N - 4) + 1]}{365^N}$$

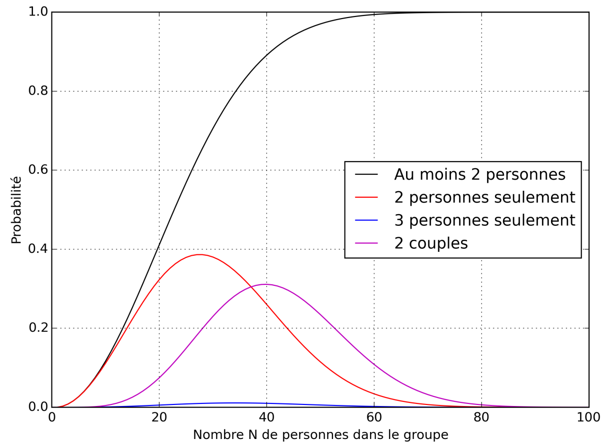

FIGURE 3 – Probabilité d'avoir au moins deux personnes ayant leur anniversaire le même jour en fonction de la taille du groupe (noir). Ainsi que la probabilité d'avoir seulement deux personnes (rouge) et seulement 3 personnes (bleu) ayant leur anniversaire le même jour. La courbe magenta est la probabilité que deux couples seulement aient leurs anniversaires deux jours distincts.

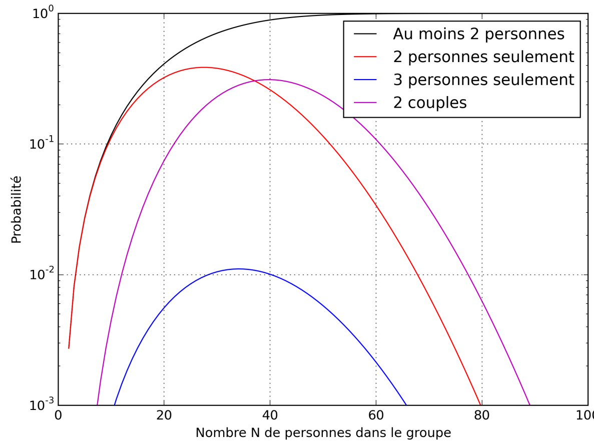

FIGURE 4 – Même chose que sur la figure 3 mais avec un axe des probablités logarithmique.

ou encore :

$$P_{2^2}^{365}(N) = \binom{365}{2} \times \binom{N}{2} \times \binom{N-2}{2} \times \frac{A_{365-2}^{N-4}}{365^N}$$

où :

- $\binom{365}{2}$ est le nombre de choix de 2 dates parmi 365 pour les deux anniversaires des deux couples ;
- $\binom{N}{2}$ est le nombre de choix des personnes pour le premier couple, soit 2 parmi N ;
- $\binom{N-2}{2}$ est le nombre de choix des personnes pour le second couple, soit 2 personnes parmi les $N-2$ restantes ;
- A_{365-2}^{N-4} est le nombre de choix possibles de dates anniversaires (sur $365-2$ dates, car 2 sont déjà prises par les deux couples) pour les $N-4$ personnes restantes ; il s'agit du nombre d'arrangements de $N-4$ parmi $365-2$.

Pour K couples, on a :

$$P_{2^K}^{365}(N) = \binom{365}{K} \times \binom{N}{2} \times \binom{N-2}{2} \times \ldots \times \binom{N-2K+2}{2} \times \frac{A_{365-K}^{N-2K}}{365^N}$$

Sachant que :

$$\begin{aligned}
\binom{N}{2} \times \binom{N-2}{2} \times \ldots \times \binom{N-2K+2}{2} &= \frac{N!}{(N-2)!2!} \times \frac{(N-2)!}{(N-4)!2!} \times \ldots \times \frac{(N-2K+2)!}{(N-2K)!2!} \\
&= \frac{1}{2^K} \times \frac{N!}{(N-2K)!} \\
&= \frac{1}{2^K} \times A_N^{2K}
\end{aligned}$$

Donc, finalement :

$$P_{2^K}^{365}(N) = \binom{365}{K} \times \frac{1}{2^K} \times A_N^{2K} \times \frac{A_{365-K}^{N-2K}}{365^N}$$

6.2.2 Généralisation

La probabilité d'un ensemble K de G personnes ayant, pour chaque groupe G, son anniversaire le même jour :

$$P_{G^K}^{365}(N) = \binom{365}{K} \times \frac{1}{G^K} \times A_N^{G\cdot K} \times \frac{A_{365-K}^{N-G\cdot K}}{365^N} \tag{3}$$

6.2.3 Formule générale

On doit donc avoir :

$$P_{\geq 2}^{365}(N) = \sum_{G=2}^{N/2} \sum_{K=1}^{N/G} P_{G^K}^{365}(N) + \sum_{J=3}^{N} P_J^{365}(N) \tag{4}$$

où le premier terme (voir Eq. (3)) est une somme sur tous les groupements de G personnes que l'on peut envisager dans le groupe de façon à en avoir au moins 2. Pour $G = 2$, on a des couples. On somme également sur tous les groupements de G personnes possibles dans le groupe (voir Eq. (**??**)). Le deuxième terme concerne les groupements de J personnes isolés.

On conçoit ainsi aisément qu'il est bien plus simple de la calculer comme nous l'avons fait (Eq. (1)) dans le paragraphe 4.

Références

[1] « La zététique ou l'art du doute : cours d'autodéfense intellectuelle, » Frédéric Baudin, 2017, cours de L3, parcours SEM à l'université Paris Sud.

[2] « Système 1, système 2 : les deux vitesses de la pensée, » Daniel Kahneman, 2011.

[3] Math Forum "Three Share a Birthday" - `http://mathforum.org/library/drmath/view/56650.html`. Consulté le 3 juillet 2017.