Trois remarques un peu bavardes sur le pari des mois d'anniversaires[1]

(n'ayant pu figurer sur l'article initial pour des questions de délais)

1) Généralisation

a) Une 'généralisation' simplette

Désignant dans le programme Python par N le nombre de parties considérées d'une année pour observer des coïndences de temps de naissance (pour les mois d'anniv N = 12, pour les dates N = 365), je constate que n_min = le nombre minimum de personnes qu'il faut dans le groupe pour avoir un pari de coïndence 'gagnant' est, évidemment !, une fonction croissante de N, cependant que n_min/N semble être une fonction décroissante de N. Je ne sais pas le prouver mais on va le 'vérifier' ensuite, je donne quelques exemples obtenus avec ce programme :

N	12	52	365	365*24	365*24*60
explicitation	Même mois de naissance	« Même semaine de naissance »	Même date d'anniv	Nés 'à la même heure'	Nés durant la même minute
n_min	5	9	23	111	854
n_min / N	0,42	0,17	0,06	0.0127	0.0016

Bien sûr comprendre que par exemple 'Nés à la même heure', 'pari gagnant' dans un groupe de 111 personnes, signifie que dans un tel groupe vous trouverez, 'probablement', deux personnes nées par exemple le 13 Septembre entre 3 et 4 heures, mais peut être en 1950 et en 2021.
Par ailleurs la formulation « Même semaine de naissance » n'est claire que si on dit 'du 1 au 7 Janvier c'est semaine 1 etc …

b) La vraie généralisation

Proposée sur la page https://fr.wikipedia.org/wiki/Paradoxe_des_anniversaires qui est très riche.

Le cadre est le suivant. Dans un ensemble à N éléments, $N > 1$, on effectue n tirages avec remise, où $1 < n < N + 1$. On désigne par $p_N(n) = p(n)$ la probabilité d'effectuer au moins deux tirages identiques.

On a vu avec $N = 12$ et $N = 365$ que $p(n) = 1 - \frac{N(N-1)\ldots(N-n+1)}{N^n}$.

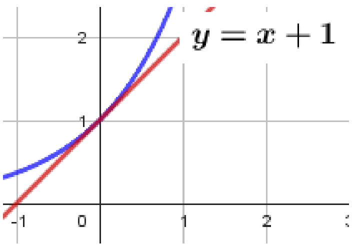
On peut préférer parler de n boules placées dans N boites.

2) Une approximation de p(n)

Sur cette page Wikipédia on trouve, expliquée avec $N = 365$, l'approximation suivante

$$p(n) \approx \quad 1 - e^{-\frac{n(n-1)}{N}} \quad \text{car } p(n) = 1 - \prod_{k=0}^{k=n-1}\left(1 + \frac{-k}{N}\right) \approx 1 - \prod_{k=0}^{k=n-1}\left(e^{\frac{-k}{N}}\right)$$

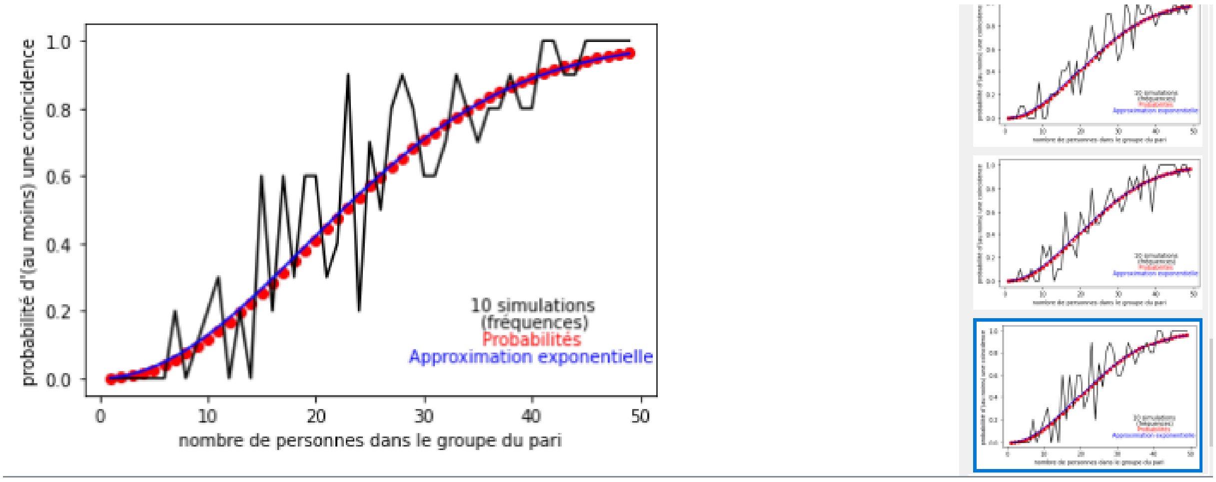
Ils parlent de développement limité d'ordre 1 de exp en 0 et effectivement … En classe de première on ne parlerait que de tangente pour dire que si $x \approx 0, e^x \approx x + 1$ … avec quand même un gros problème pour expliquer qu'on peut tranquillement multiplier n telles approximations …



$$\text{Finalement } p(n) \approx 1 - e^{\sum_{k=1}^{k=n-1}\left(\frac{-k}{N}\right)} = 1 - e^{-\frac{n(n-1)}{2N}}.$$

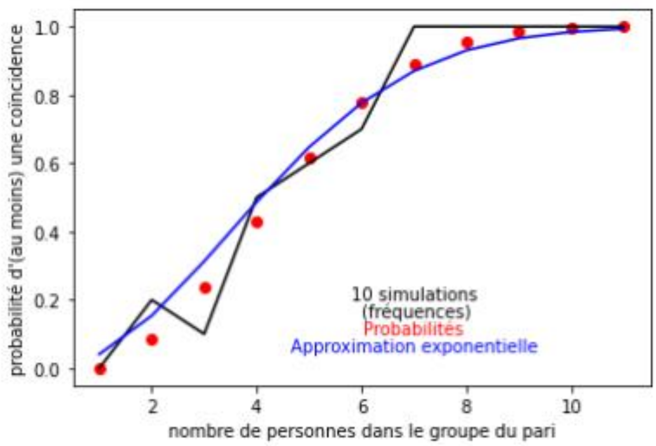
[1] J Kentzel Auch

Pour le pari des annivs, $N = 365$, l'approximation est très convenable pour n assez petit devant N, ci-dessous $n < 50$.



(en passant ces dessins, montrant les exécutions précédentes, viennent de Python avec Spyder)

Pour le pari des mois d'annivs, l'approximation demeure correcte alors que $N = 12$ est petit !



Sur la page Wikipédia ils approximent encore $p(n)$ par $p(n) \approx \; 1 - e^{-\frac{n^2}{2N}}$ et du coup peuvent, pour un entier fixé N, pour une probabilité p donnée, obtenir n dans $p = 1 - e^{-\frac{n^2}{2N}}$: n est le 'nombre' de membres du groupe qu'il faut pour gagner le pari des anniversaires (en coupant une année en N $parties$) avec probabilité p : $n = \sqrt{2N.Ln\left(\frac{1}{1-p}\right)}$.

On trouve n_min, évoqué au 1)a), le nombre minimum de personnes qu'il faut dans le groupe pour avoir un pari de coïndence 'gagnant' en prenant $p = 0{,}5$. On obtient, valeur par défaut car en fait on voulait $p > 0{,}5$, $n_{min} = \sqrt{2N.Ln\,2}$. Comme évoqué au 1)a) on a bien f définie par $f(N) = \frac{n_{min}}{N}$ qui est décroissante. On vérifie les cinq valeurs qui avaient été laborieusement calculées avec Python dans le tableau du début :

```python
# -*- coding: utf-8 -*-

from math import sqrt,log,exp

def n_min(N):
    return sqrt(2*N*log(2,exp(1)))
    # pour avoir Ln(x) avec Python, taper log(x,e)

for n in [12,52,365,365*24,365*24*60]:
    print(n,',',round(n_min(n),3))
```

On obtient cette sortie :

```
12 , 4.079
52 , 8.49
365 , 22.494
8760 , 110.2
525600 , 853.602
```

En passant on imagine que c'est cette formule d'approximation, qui peut être présentée rigoureusement, qui a permis à Richard von Mises, décédé en 1953, il y avait alors bien peu d'ordinateurs pour calculer 365*364* …, de parler de ce paradoxe des anniversaires.

3) Une formule semblant compliquée mais évidente pour nous

N est un entier, valant au moins 2.

Pour tout entier n vérifiant $2 \leq n \leq N+1$, on a l'égalité

$$(\mathrm{N}-1)! * \sum_{k=1}^{n-1} \frac{k}{N^k*(N-k)!} = 1 - \frac{N!}{(N-n)!*N^n} \quad .(1)$$

On peut préférer l'écrire ainsi :

$$\sum_{k=1}^{n-1} \frac{k(N-1)(N-2)\dots(N-k+1)}{N^k} = 1 - \frac{N(N-1)\dots(N-n+1)}{N^n} \quad (2),$$

mais alors le terme de la somme de gauche n'a plus de sens pour k = 1 !,

en effet la « sorte d'identité » $\frac{a!}{b!} = a(a-1)\dots(b+1)$ n'est correcte que si $a > b$,

cette forme serait correcte :

$$\frac{1}{n} + \sum_{k=2}^{n-1} \frac{k(N-1)(N-2)\dots(N-k+1)}{N^k} = 1 - \frac{N(N-1)\dots(N-n+1)}{N^n}.$$

On pourrait aussi la déguiser ainsi pour la rendre redoutable en tentant de cacher le 1 qui montre qu'elle vient du calcul des probas :

$$\sum_{k=1}^{n-1} \frac{k}{N^k*(N-k)!} = \frac{1}{N^n(N-1)!}\left(N^n - \frac{N!}{(N-n)!}\right) \quad (3),$$

ou, pire exercice : calculer les entiers a, b, c, d et e vérifiant $(\mathrm{N}-1)! * \sum_{k=1}^{n-1} \frac{k}{N^k*(N-k)!} = a + \frac{b\,!}{c\,!*d^e}$.

(1) peut être facilement prouvée par récurrence, finie, sur n, avec n entre 2 et N+1 :

$\sum_{k=1}^{n} u_k = \sum_{k=1}^{n-1} u_k + u_n = \dots$.

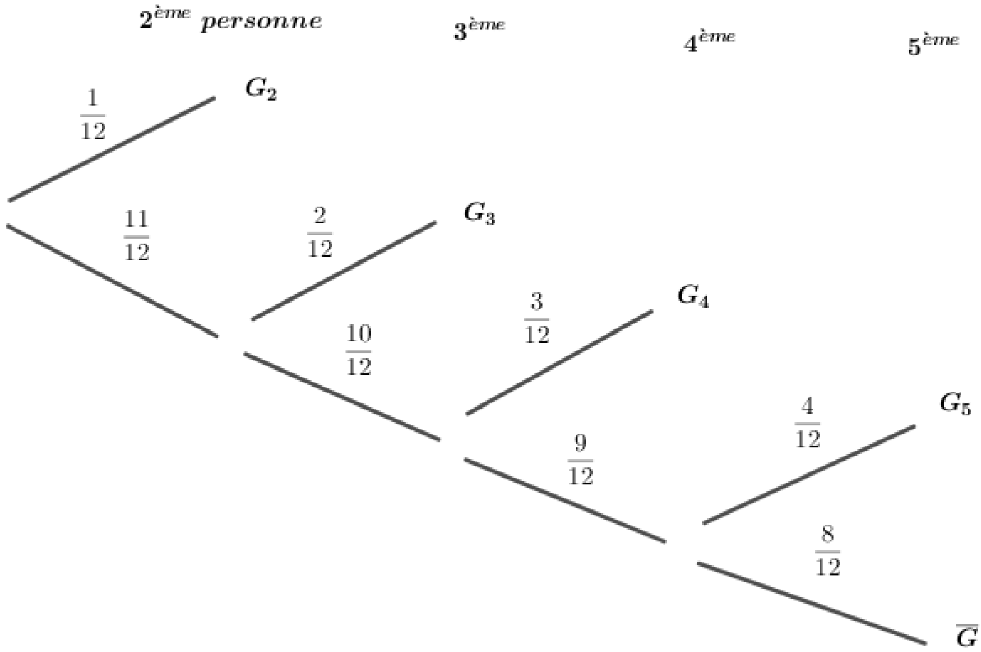
On l'écrit, sous la forme (1), avec N = 12 et n = 5 pour voir qu'il ne s'agit que de deux façons d'expliquer pourquoi avec 5 personnes le pari des mois d'anniv est 'gagnant' :

$$\frac{1}{12}. + \frac{2*11}{12^2} + \frac{3*11*10}{12^3} + \frac{4*11*10*9}{12^4} = 1 - \frac{12*11*10*9*8}{12^5}$$

On pose G = {gagner ce pari}

Façon 1 : avant la classe de première, on ne connait que les arbres de dénombrement, c'est le membre de droite.

Façon 2 : on connait les probabilités conditionnelles et les arbres de probabilités, les gens écrivent, chacun à son tour, leur mois de naissance sur la même feuille, il y a 4 instants où le pari peut être gagné. G est la réunion disjointe des événements G_2, G_3, G_4 et G_5.



En passant j'avoue que j'ai l'impression que je n'avais jamais dessiné en classe ce petit arbre [2], car j'ai toujours considéré cette histoire des anniversaires comme la plus fameuse utilisation en calcul des probas de la bête formule du complémentaire et j'ai toujours orienté les élèves dans cette direction. La routine …

Ce n'est qu'en essayant à nouveau de me passer de cette formule, pour me persuader qu'elle est incontournable (mais pour moi avec seulement des dénombrements c'est inextricable), que j'ai croisé cet arbre !

Si vous posez la question des <u>mois</u> d'anniversaire sans orienter les élèves vous avez surement des chances de le voir apparaitre …

[2] Une sorte de loi géométrique, temps d'attente ou d'arrivée du premier succès, mais avec la probabilité du succès qui augmente.