Université de Liège

Ingénieur Civil - Bachelier

Projet 1

Paradoxe des anniversaires

MUKOLONGA Jean-David 20170679
SIBOYABASORE Cedric 20175202

17 mars 2019

1 Question 1 : Méthode de Lehman Leighton

1.a

En utilisant la fonction ***naiveBirthdaySol*** pour deux personnes , nous obtenons une probabilité de 0.0027. La réponse nous paraît correcte car en calculant nous-même cette probabilité,en effectuant 365 x $\frac{1}{365}$ x $\frac{1}{365}$, nous tombons sur la réponse fournie par la fonction.
Les fonctions ***tic*** et **toc** nous révèlent que MATLAB a pris 1.859382 secondes pour effectuer ce calcul.

1.b

La fonction ***naiveBirthdaySol*** avec $N = 3$ nous prend énormement de temps.
En supposant que le temps d'exécution t est proportionnel au nombre de branches de l'arbre construit par ***createBirthdayTree*** tel que

$$t = \alpha\ 365^N \text{ où } \alpha \text{ est une constante à déterminer,}$$

nous trouvons $\alpha = 1.3957\text{e-}05$ grâce aux conditions imposées par $t = 1.859382$ secondes pour $N = 2$.
En ayant déterminé α, nous pouvons évaluer les temps d'éxecution de ***naiveBirthdaySol*** pour différentes valeurs de N (voir Tableau 1).

N	Temps estimé (en secondes)
3	678.67443 soit 11.3 minutes
4	247716.167 soit près de 3 jours
5	90416400.94 soit près de 3 ans

Tableau 1

En utilisant cette approche, nous remarquons que la fonction ***naiveBirthdaySol*** prend énormément de temps pour des valeurs de N supérieures à 2.

1.c

Notons A l'évènement décrit par la condition "il y a au moins 2 personnes parmi N nées le même jour".
Le contraire d'un tel évènement est l'évènement $\bar{A}$ défini par "aucune des N personnes n'est née le même jour".

Calculer la probabilité de l'évènement contraire $\bar{A}$ est plus simple car nous considérons uniquement les branches de l'arbre pour lesquelles les N personnes sont nées un jour différent. Nous pourrons ensuite déterminer la probabilité de A.
Pour $N = 3$, nous avons un arbre tel qu'affiché sur la Figure 1.

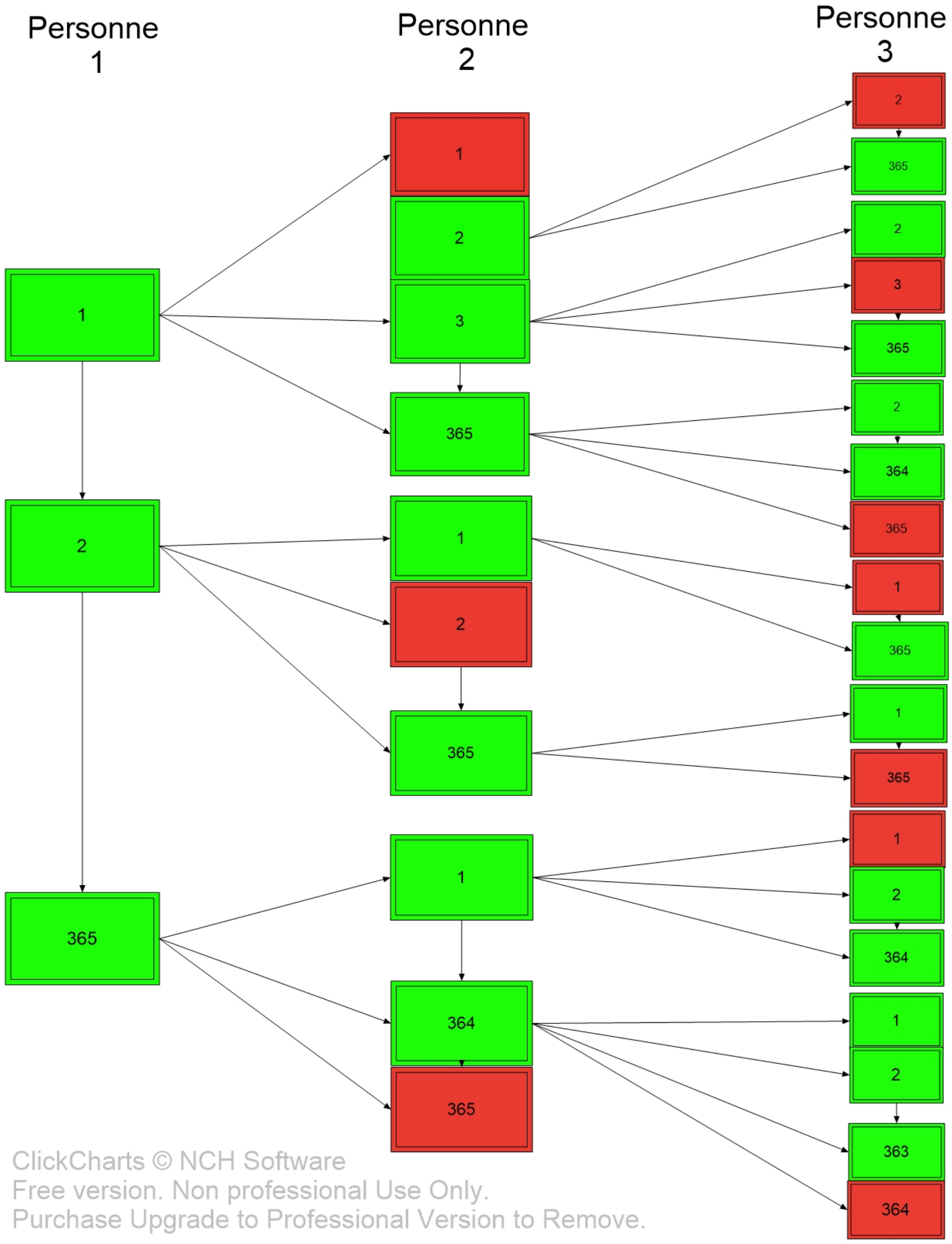

Figure 1

La personne 1 a une probabilité de $\frac{1}{365}$ d'être née un certain jour de l'année. Or, il y a 365 jours différents dans l'année. Elle a donc une probabilité de $\frac{365}{365}$ d'être née un jour de l'année. Sur la Figure 1, nous représentons un jour de l'année par une case, le 1 janvier par 1 et le 21 décembre par 2 etnous colorions en vert toutes les cases de la personne 1. Nous représentons par une longue ligne verticale les jours entre le 2 janvier et 31 décembre pour que l'arbre ne prenne pas trop de place.

La personne 2 a une probabilité de $\frac{365}{365}$ d'être née un jour de l'année. Or, comme nous ne considérons que les branches correspondant à l'évènement $\bar{A}$, nous éliminons le cas où la personne 2 est née le même jour que la personne 1 (nous colorions en rouge les cases correspondant à ce cas et en vert les cases correspondant au cas contraire). La personne 2 a donc une probabilité de $\frac{364}{365}$ d'être née un jour différent de la personne 1.

La personne 3 ne peut pas être née le même jour que la personne 1 et que la personne 2. Elle a donc une probabilité de $\frac{363}{365}$ d'être née un jour différent des deux premières personnes.

Pour $N = 3$, la probabilité associée à $\bar{A}$ vaut donc $\frac{365}{365} * \frac{364}{365} * \frac{363}{365} = 0.9917958341$.
Par complémentarité, nous pouvons calculer la probalité de A qui vaut
1 - $\frac{365}{365} * \frac{364}{365} * \frac{363}{365} = 0.0082$.

Nous pouvons tirer de ce raisonnement que la probabilité qu'au moins deux personnes parmi un groupe de N personnes soient nées le même jour est donnée par

$$\mathbf{1} \text{ - } \frac{365-(N-1)}{365^N}$$

Le résultat se rapprochera de plus en plus de 1 lorsque N augmente.

1.d

Notre fonction ***Q1d*** se base sur notre arbre simplifié. Elle prend en entrée le nombre de personnes N du groupe et retourne la probabilité qu'au moins deux personnes de ce groupe aient leurs anniversaires le même jour.
Utilisons la pour déterminer les probabilités pour $N = 2, 3, 4, 5, 20, 30, 40, 50, 60, 80$. que nous listons dans le Tableau 2

N	Probabilité	Temps mesuré	Temps estimé 1b)
2	0.0027	0.000006	1.8594
3	0.0082	0.000007	678.6744
4	0.0164	0.000008	247716.167
5	0.0271	0.000006	90416400.94
20	0.4114	0.000007	2.4583e+46
30	0.7063	0.000007	1.0317e+72
40	0.8912	0.000008	4.3301e+97
50	0.9704	0.000009	1.8173e+123
60	0.9941	0.000010	7.6270e+148
80	0.9999	0.000009	1.3434e+200

Tableau 2

Nous pouvons constater que les temps mesurés (en secondes) avec la fonction ***Q1d*** sont largement inférieurs aux temps estimés (également en secondes) pour ***naiveBirthdaySol***.
Au vu des probabilités obtenues, nous pouvons conclure que plus le groupe s'agrandit, plus il y a de chances qu'au moins deux personnes aient leurs anniversaires le même jour.
Grâce à ***Q1d***, nous pouvons conclure qu'il faut un groupe de $N = 23$ personnes pour avoir 50% de chances qu'au moins deux personnes issues de ce groupe soient nées le même jour et il faut un groupe de $N = 57$ personnes pour avoir 99% de chances qu'au moins deux de ces N personnes soient nées le même jour.

2 Question 2

2.a

Cette variable X suit une loi de Bernoulli. En utilisant la fonction créée au point 1.d, nous pouvons aisément trouver la probabilité du succès et donc de l'échec. L'espérance E nous est donnée par la probabilité du succès p et la variance V nous est fournie par la formule suivante

$$V = p(1 - p)$$

Nous obtenons donc $E = 0,8782$ et $V = 0,1069$.

2.b

En comparant les résultats obtenus avec le script ***MonteCarlo.m*** et les valeurs théoriques, nous nous apercevons que la méthode de Monte-Carlo

nous donne une bonne approximation de l'espérance et de la variance.
En augmentant la taille du vecteur, l'approximation tend vers la valeur théorique. Toutefois, plus nous augmentons la taille des vecteurs, plus l'écart entre eux diminue ; il n'est donc pas utile d'augmenter la taille indéfiniment mais il faut trouver un juste milieu entre le temps de calcul de la machine et la précision de l'approximation souhaitée.

2.c

L'observation du Tableau 3 nous permet de retirer deux choses :
Premièrement, pour les espérances, lorsque la taille des vecteurs augmente, les approximations se rapprochent de la valeur théorique.
Deuxièmement, les variances diminuent de plus en plus quand la taille du vecteur augmente. Ceci paraît logique car la valeur dans le vecteur se rapprochent de plus en plus lorsque la taille de ce dernier augmente.

taille des vecteurs	10	10^2	10^3	10^4
moyenne de l'espérance	0.8778	0.8790	0.8782	0.8782
variance de l'espérance	0.0117	0.0010	0.0001	0.0000
moyenne de la variance	0.0956	0.1053	0.1068	0.1070
variance de la variance	0.0054	0.0006	0.0001	0.0000

Tableau 3

2.d

Lorsque la taille du vecteur augmente, la valeur de la moyenne et la valeur de l'espérance ne tendent pas vers des valeurs précises.
De plus, en répétant l'opération plusieurs fois, nous nous aperçevons que les nombres différent d'essai en essai (voir TEST 1, TEST 2 et TEST 3).La question du pourquoi s'impose. La réponse réside dans le type de distribution que suit la variable aléatoire Y.
En effet, la variable Y suit une distribution dite de Cauchy. Cette loi n'admet ni espérance ni variance. Ainsi, les nombres calculés n'ont aucune signification physique.

TEST 1 :

taille des vecteurs	10	10^2	10^3	10^4
moyenne des espérances	-8.0837e-04	5.8616e-04	0.0030	-0.0012
variance des espérances	6.5281e-04	3.4324e-04	0.0093	0.0015

TEST 2 :

taille des vecteurs	10	10^2	10^3	10^4
moyenne des espérances	4.6595e-04	7.1520e-04	0.0012	-0.0097
variance des espérances	2.1689e-04	5.1100e-04	0.0014	0.0932

TEST 3 :

taille des vecteurs	10	10^2	10^3	10^4
moyenne des espérances	-0.0035	6.7214e-04	3.5972e-04	0.0016
variance des espérances	0.0121	4.5131e-04	1.2927e-04	0.0026