

Première approche de la simulation en première L

Bernard EGGER

Les élèves ont eu un mois et demi environ de familiarisation avec le tableur et connaissent déjà un certain nombre de fonctions de base. La démarche n'est donc sans doute pas exportable sans adaptation à une classe de seconde pour laquelle l'utilisation d'un tableur a une place plus marginale. Il s'agit de les amener en quelques séances à comprendre les objectifs de la simulation sur tableur et sa mise en place effective dans un certain nombre de situations simples.

1 Pour commencer : étude d'une liste de nombres

Une fichier contenant une liste de 100 chiffres est proposé. La question est la suivante : combien y a-t'il de « 0 », de « 1 », de « 2 »..., de « 9 » dans cette liste. La liste est identique pour tous les élèves, elle est protégée pour qu'on ne puisse pas la modifier.

Pour 100 nombres il est possible de les compter mais ce serait certainement plus facile sur une feuille que sur un ordinateur. Il s'agit donc de trouver une méthode alternative basée sur la façon dont on procède pour compter. Si par exemple, pour savoir combien il y a de 0, on ajoute 1 chaque fois que l'on rencontre un " 0 " et l'on ne compte pas les autres chiffres (ce qui revient à ajouter 0). Le codage est simple : dans la colonne B, mettons 1 chaque fois que l'on rencontre un 0 dans la colonne A, et 0 chaque fois que l'on rencontre un autre chiffre ; il ne reste plus qu'à faire la somme de la colonne B, ce qui se fait simplement sur Excel. Toutefois cela reste très fastidieux (les élèves ont fait d'abord cela « à la main »). Il faudrait pouvoir disposer d'une fonction qui automatiserait la procédure décrite ci-dessus. Au tableau est écrit :

$$\text{Si } A_i = 0 \text{ alors } B_i = 1 \text{ sinon } B_i = 0$$

Il est alors facile d'introduire la fonction **Si** d'Excel : **Si** ($A_i = 0 ; 1 ; 0$). Le plus important est bien sûr de remarquer que vis à vis du test, tous les nombres autres que « 0 » jouent le même rôle.

Les élèves sont ainsi invités à « compter » d'autres chiffres.

La deuxième question posée est de « compter » combien il y a de nombres pairs. Il faudrait pouvoir écrire :

$$\text{Si } (A_i = \text{pair} ; 1 ; 0)$$

Les élèves cherchent si Excel possède la fonction **pair**. Mais déception, elle ne dit pas si un nombre est pair. Pour comprendre comment elle fonctionne on peut examiner son action sur les premiers termes de la liste :

7	8
5	6
5	6
2	2
3	4
7	8
0	0
7	8
8	8
7	8
0	0
4	4

Il semble que chaque fois que l'on a un nombre impair, Excel rajoute 1, par contre les nombres pairs sont inchangés. Comment utiliser ce résultat pour mettre en place le test ? Il suffit bien sûr de tester si un nombre est transformé ou pas par exemple en étudiant la valeur de la différence entre deux cellules A_i B_i . Le test " Pair " est alors :

$$\text{Si } (B_i - A_i = 0 ; 1 ; 0)$$

ou simplement :

$$\text{Si } (A_i = B_i ; 1 ; 0)$$

La troisième question abordée est plus délicate : peut-on savoir combien de nombres de la liste sont divisibles par 3 ?

7	2,33333333
5	1,66666667
5	1,66666667
2	0,66666667
3	1
7	2,33333333
0	0
7	2,33333333
8	2,66666667
7	2,33333333
0	0

Ici peut d'espérer d'avoir une fonction « trois » dans Excel. Tout d'abord les élèves sont invités à « diviser par 3 » chaque terme de la liste.

La division tombe juste si le quotient est un nombre entier. Il faudrait donc pouvoir tester si un nombre est entier, avec quelque chose du type :

$$\text{Si } (A_i \text{ entier} ; 1; 0)$$

Dans leur recherche sur les fonctions disponibles sur Excel, les élèves ne trouvent pas de fonction « Entier », mais quelque chose qui y ressemble : **Ent.** (la partie entière) qu'ils sont invités à tester sur la colonne B ci-dessous :

7	2,33333333	2
5	1,66666667	1
5	1,66666667	1
2	0,66666667	0
3	1	1
7	2,33333333	2
0	0	0
7	2,33333333	2
8	2,66666667	2
7	2,33333333	2
0	0	1

Il est immédiat que si $B_i = C_i$, le nombre contenu dans A_i est divisible par 3, sinon il ne l'est pas. On peut donc construire un test :

$$\text{Si } (B_i = C_i; 1; 0)$$

Cette façon de procéder est coûteuse. On peut la résumer en B (mais cela passe un peu moins bien auprès des élèves) :

$$\text{Si } (A_i/3 = \text{ent}(A_i/3); 1; 0)$$

Quelle interprétation peut-on faire de la colonne C ? A quoi correspond la partie entière du quotient de A_i par 3 ? Petit débat, puis rapidement la bonne idée : il s'agit du quotient dans la « division entière » de A_i par 3. Du coup on peut facilement calculer un autre élément important de cette division : le reste.

7	2,33333333	2	1
5	1,66666667	1	2
5	1,66666667	1	2
2	0,66666667	0	2
3	1	1	0
7	2,33333333	2	1
0	0	0	0
7	2,33333333	2	1
8	2,66666667	2	2
7	2,33333333	2	1
0	0	1	0

Cela met en évidence un nouveau test pour la divisibilité par 3 :

$$\text{Si } (D_i = 0; 1; 0)$$

Mais aussi une « évidence » : les restes possibles dans la division par 3 sont 0, 1 et 2 (ce que les élèves comprennent fort bien).

On peut donc trouver trois catégories de nombres en fonction de leur reste et en déterminer l'effectif pour chaque reste.

On reprend l'exercice avec la division par 4.

2 Pour poursuivre : à la découverte des nombres aléatoires

La partie précédente a permis un premier contact avec des opérations sur une liste donnée de nombres entiers. Il s'agit dans un deuxième temps de construire des listes de nombres « au hasard ».

0,51023229
0,32702429
0,91684876
0,58296496
0,50993949

L'approche la plus simple est bien sûr la fonction **Aleat.**

Les élèves sont très surpris par cette « fonction » qui ne retourne pas toujours le même résultat.

Ils constatent que toutes les réponses sont entre 0 et 1 (non compris).

Comment faire pour obtenir des nombres aléatoires compris entre 0 et 10 non compris? Il suffit de multiplier les précédents par 10. Et comment faire pour avoir des nombres aléatoires entiers compris entre 0 et 10? Il suffit de prendre la partie entière de la colonne précédente.

0,85639863	8,56398631	8
0,66932053	6,69320531	6
0,60825777	6,08257771	6
0,91089452	9,10894516	9
0,10636850	1,06368503	1

Tout se passe bien à un petit détail près que nous examinerons un peu plus loin.

On peut « automatiser » le calcul :

$$\text{ent}(A_i * 10)$$

Ainsi nous sommes capables d'engendrer des nombres entiers au hasard compris entre 0 et 9.

Regardons de plus près le « petit détail » : à chaque nouvelle action, Excel procède à un recalcul. Ce qui a pour effet automatique de changer les valeurs rendues par la fonction **Alea**. Nous avons bien une suite de nombres mais elle n'est pas stable, donc peu intéressante à étudier. Bloquer le recalcul n'est pas très utile : dans ce cas on perd la recopie automatique ... Il reste un moyen que l'on étudie avec les élèves : la programmation (modeste) d'une fonction hasard.

```
Function hasard() hasard = Rnd() End Function
```

Et là tout fonctionne de façon beaucoup plus stable.

Nous avons été capable de trouver des nombres entiers au hasard compris entre 0 et 9, mais pourrait-on de même en trouver entre 0 et 5, entre 0 et 6, entre 1 et 3 ...

3 Pour terminer : première approche de la simulation

Reprenons les questions sur lesquelles nous avons fini la partie précédente. Comment trouver des chiffres au hasard compris entre 0 et 5? Pour obtenir des chiffres au hasard compris entre 0 et 9, nous avons multiplié les résultats rendus par la fonction *hasard* par 10. Donc ici on peut supposer que l'on obtiendra des chiffres au hasard compris entre 0 et 5 en multipliant les résultats de *hasard* par 6. Puis on prendra la partie entière.

0,15566307	0,93397844	0
0,47445917	2,84675503	2
0,25726765	1,54360592	1
0,62875187	3,77251124	3
0,54207021	3,25242126	3
0,15630221	0,93781328	0
0,93854517	5,631271	5
0,65449941	3,92699647	3
0,50608736	3,03652418	3

Cela semble marcher.

On peut le prouver de façon presque élémentaire :

$$0 \leq x < 1 \text{ donne } 0 \leq 6x < 6$$

et donc : $0 \leq \text{ent}(x) < 6$.

Autrement dit les seules valeurs possibles sont 0, 1, 2, 3, 4, 5.

On peut alors construire des nombres au hasard compris entre 1 et 6. Il suffit bien sûr d'ajouter 1 au précédent. On peut alors créer une fonction qui donnera directement ce résultat :

```
Function hasard6() hasard6 = Int(6 * Rnd()) + 1 End Function
```

A quoi pourrait-on assimiler les résultats obtenus? Les élèves le disent vite : à ce que l'on obtient en jetant un dé cubique. Mais quelle condition doivent-ils remplir pour correspondre à cette situation? Bien évidemment ce qui est évoqué ici c'est l'équiprobabilité. Et nous avons maintenant les outils pour tester cette équiprobabilité. On procède à la constitution d'une liste de 600 nombres (imaginez qu'il faille jeter 600 fois un dé ...).

On pourrait obtenir par exemple :

598	2	0	1	0	0	0	0
599	3	0	0	1	0	0	0
600	4	0	0	0	1	0	0
		85	114	117	101	104	79

Chaque élève obtiendra un résultat différent (la fluctuation d'échantillonnage). On peut bien sûr faire la moyenne de ces résultats ... Mais on peut aussi prendre tout simplement un dé (par élève) et se

répartir un nombre de jets pour chacun permettant d'obtenir au moins 600 résultats. La fluctuation d'échantillonnage toujours à l'œuvre, les résultats obtenus ressembleront assez probablement à ceux du tableur.

Si le tableur est un bon moyen de simuler le lancer d'un dé, c'est parce que nous avons construit un modèle mathématique conforme à celui auquel nous pensons que le dé obéit : la situation d'équiprobabilité. Mais aussi parce que le générateur de nombres aléatoires d'Excel est assez performant quant à l'équiprobabilité ...

4 Quelques remarques

Commençons par une première remarque qui vient immédiatement à l'esprit : comment être sûr que le premier chiffre de la partie décimale des nombres aléatoires d'Excel suit bien une loi d'équiprobabilité. De cela nous ne serons jamais sûrs bien entendu. Tout au plus pouvons nous le tester sur un très grand nombre de tirages de nombres aléatoires. Il faut essayer, mais cela semble bien se confirmer ...

On pouvait craindre en particulier que la répartition suive plutôt une loi du type loi de Benford. C'est une loi très surprenante que cette loi de Benford. Si l'on prend une série de nombres comme par exemple les cours des actions boursières, les prix dans un supermarché, les longueurs des fleuves du globe et que l'on mesure les fréquences du chiffre le plus à gauche différent de 0 (autrement dit le premier chiffre significatif), les neuf chiffres de 1 à 9 n'apparaissent pas avec la même probabilité. Essayez si vous n'êtes pas convaincu !

La loi de Benford affirme que la probabilité qu'un nombre tiré d'un ensemble quelconque comporte un premier chiffre significatif égal à d est donné par la formule suivante:

$$\Pr(\text{« le 1er chiffre significatif est } d \text{ »}) = \log_{10} \left(1 + \frac{1}{d}\right), \text{ avec } d = 1, 2, \dots, 9$$

Est-on vraiment dans ce cas pour le chiffre le plus à gauche de la partie décimale d'un nombre aléatoire d'Excel? En dehors du fait que « 0 » est aussi possible, on peut penser que oui. Un test peut être réalisé sur Excel.

On construit d'abord une fonction qui pour chaque nombre aléatoire d'Excel rend le premier chiffre différent de 0 en partant de la gauche.

Le programme de la fonction **benford** est présenté ci-dessous.

Le tableau montre, pour une liste de 3600 chiffres placés en colonne A, les résultats obtenus.

A	B	C	D	E	F
7		1	357	400	1083,70798
5		2	405	400	633,928533
5		3	431	400	449,779452
2		4	409	400	348,876047
3		5	398	400	285,052486
7		6	411	400	241,008443
1		7	384	400	208,771009
7		8	402	400	184,149081
8		9	399	400	164,726966

```
Function benford()
    i=0
    j=0
    Do While i=0
        k=10*j
        i=Int(k)
        j=k-i
    Loop
    benford = i
End Function
```

La colonne A contient comme cela a été dit 3600 chiffres obtenus par la fonction benford. La colonne C les chiffres de 1 à 9. La colonne D contient pour chaque chiffre de C le nombre d'apparitions de ce chiffre dans la liste A. Elle a été construite en utilisant la fonction d'Excel : NB.SI. La syntaxe est par exemple pour D1 :

=NB.SI(A1 : A3600 ; C1)

La colonne E ne contient que $3600/9 = 400$, référence d'une répartition régulière. La colonne F contient la répartition selon la loi de Benford : ainsi, pour F1 :

=log10(1+1/C1)*3600

Même si l'on ne tient pas compte des zéros, et en prenant les précautions d'usage quant à l'inévitable fluctuation d'échantillonnage, il semble bien que les résultats obtenus soient plus près d'une distribution régulière que d'une distribution selon la loi de Benford.

Sites sur la loi de Benford :

http://plvw.free.fr/sam_benford.html

<http://www.jura.ch/lcp/cours/dm/benford/benford.html>

http://www.inrialpes.fr/sel/articles/benford/cadre_benford.html

<http://www.larecherche.fr/VIEW/316/03160721.html>

<http://www.arte-tv.com/hebdo/archimed/19981229/ftext/sujet7.html>