LES GENERATEURS DE NOMBRES ALEATOIRES

Ce travail a deux objectifs :

==

1. Comprendre ce que font les générateurs de nombres aléatoires RAND, ALEA() ...
2. En tirer trois sujets pour la nouvelle épreuve pratique de mathématiques en TS

==

1. La fonction RAND de la calculatrice (ou ALEA() du tableur EXEL) donne un « nombre aléatoire » $u_n \in [0;1[$. Pour se convaincre qu'il s'agit bien de $[0;1[$ et non de $[0;1]$ il suffit d'activer la fonction ALEA() d'EXEL, la réponse est donnée en toutes lettres.

Qu'est ce qu'une suite de nombres aléatoires $u_n \in [0;1[$?

Il n'y a pas de définition exploitable, tout juste peut-on avoir des exigences du type :

a) le premier chiffre de la partie décimale sort avec la probabilité 1/10,
les deux premiers chiffres avec la probabilité 1/100, etc ….
Ces conditions permettent d'envisager des procédures créant des chiffres 0123456789 dont les fréquences d'apparition dans tout échantillon de taille n « assez grand » avoisinent 1/10, *cf : sujets n°1 et 2 ;* par contre elles ne permettent pas d'affirmer qu'un tel échantillon est une table (ou liste) de chiffres aléatoires. Voir pour cela le test du poker avec le commentaire d'Arthur ENGEL.

b) Une telle suite $(u_n)_{n \in IN^*}$ doit être équirépartie sur $[0;1]$, en ce sens :
quel que soit $[a;b]$ inclus dans $[0;1]$, si, pour $n \in IN^*$, on note respectivement $q(n)$ et $f(n) = q(n)/n$ le nombre $q(n)$ d'éléments u_p de $\{u_1 , u_2 , ... , u_n\}$ appartenant à $[a;b]$ et leur fréquence $f(n)$,
alors la suite $(f(n))_{n \in IN^*}$ tend vers b-a
cf : sujet n°2

Remarques :

1) L'exigence (a) est réalisée en cas d'équirépartition. *cf : sujet n°2*

2) L'équirépartiton permet de définir, pour tout $[a;b]$ contenu dans $[0;1]$, la probabilité p de l'événement "$u_n \in [a;b]$" par $p = b-a$.

3) Dans la pratique, il est impossible de fabriquer une suite de nombres aléatoires infinie;
elles sont donc toutes finies (ou plutôt périodiques) et leur conception est un secret d'ingénieur.
Le sujet n°2 fournit des procédés de fabrication de chiffres aléatoires à partir de nombres aléatoires.

4) La vocation d'une suite de nombres aléatoires de $[0;1[$ est, au moins en partie, d'engendrer des tables de chiffres aléatoires.
Considérons la suite de terme général $v_n = \sqrt{n} - E(\sqrt{n})$ où $E(x)$ désigne la partie entière de x; cette suite <u>infinie</u>, bien qu'équirépartie sur $[0;1]$ au sens mathématique du terme défini plus haut, engendre des tables de chiffres de probabilité d'apparition 10^{-1} (voir le sujet n°2) par blocs de 5 par la procédure « $w_n = E(10^5.(\sqrt{n} - E(\sqrt{n})))$ » qui ne vérifient pas le "test du poker" défini plus loin pour des échantillons $\{w_1 , w_2 , ... , w_n\}$ de dimension n raisonnables (ce qui m'a fait dire, trop vite, que l'équirépartition d'une suite sur $[0;1]$ n'était pas une condition suffisante pour en faire une suite de nombres aléatoires); il y a dans la suite (v_n) des termes « parasites », les carrés d'entiers, qui font que la 7^{ieme} catégorie du test du poker est sur représentée dans ces échantillons. Ces termes « parasites » ont pour fréquence $(E(\sqrt{n})/n)$ qui tend vers 0 comme $(1 / \sqrt{n})$, c'est à dire trop lentement pour que leur influence soit négligeable pour des valeurs de n raisonnables. Peut être a t'on là une des raisons qui font que les suites infinies sont inexploitables ?
Pour plus de détails sur ce qui précède <u>voir le document « racine(n) - partie entière de racine(n) »</u>.
(la démonstration de l'équirépartition de cette suite (v_n) est accessible à un élève de TS)

LE TEST DU POKER

Citons Arthur ENGEL dans « Les certitudes du hasard » page 147 §11.2 :

"Aucun procédé de fabrication de chiffres aléatoires n'est entièrement fiable. Il est donc nécessaire de « mesurer » leur caractère aléatoire. En particulier il ne suffit pas de mesurer la fréquence de chaque chiffre il faut aussi vérifier la fréquence de différents blocs.
Un des tests les plus sûrs est le test dit du poker. Les chiffres sont regroupés par blocs de 5.
La probabilité d'un tel bloc est 10^{-5} . Ces blocs sont regroupés en 7 catégories dont les probabilités sont calculées. On compare alors ces probabilités avec les fréquences observées ".

Le tableau suivant donne la description des 7 catégories, lointainement inspirées du poker (je cite) :

n°	description	type	exemple	probabilité
1	chiffres différents	abcde	34961	0,3024
2	une paire	aabcd	29512	0,5040
3	deux paires	aabbc	44533	0,1080
4	un triplet	aaabc	60366	0,0720
5	paire triplet	aaabb	23223	0,0090
6	quadruplet	aaaab	29222	0,0045
7	quintuplet	aaaaa	55555	0,0001

Ce test est à l'origine du sujet n°3

<u>Remarque</u> : pour se convaincre que dans une liste de chiffres il ne suffit pas de vérifier que chacun des chiffres apparaît avec la probabilité 1/10 pour en faire une liste de chiffres aléatoires il suffit de considérer la liste {012345678901234567890123456789………0123456789…..}, où chacun des chiffres apparaît avec la probabilité 1/10, mais qui n'a rien d'une liste de chiffres aléatoires, ceci bien qu'aucune définition de liste de chiffres aléatoires n'ait été donnée.

SUJET N°1

SUITES DE NOMBRES ALEATOIRES ET TEST D'ADEQUATION À UNE LOI UNIFORME.

Cet exercice peut se traiter sur tableur ou sur machine

La variable aléatoire X désigne le résultat obtenu par une procédure créant un chiffre aléatoire 0, 1, 2, ... , ou 9 à partir du nombre aléatoire fourni par la fonction ALEA() du tableur (ou par la fonction RAND de la calculatrice).

Peut - on considérer qu'un résultat $i \in \{0; 1; 2; ... ; 9\}$ apparaît avec la probabilité $pi = \frac{1}{10}$?

Pour ce faire, on réalise n tirages :
les résultats obtenus lors de la répétition de ces n tirages constituent un échantillon de taille n.
On teste l'adéquation des fréquences f_i à une loi de probabilité uniforme.

Notations :
f_0 , f_1 , ... , f_9 désignent les fréquences respectives des résultats X = 0; X = 1; X = 2; ... ; X = 9 dans l'échantillon et $\{ \ pi = p(X= i) = \frac{1}{10}, \ i \in\{0; 1; 2; ... ; 9\} \ \}$ désigne la loi de probabilité de la loi uniforme ;

le test suivant est dérivé du test du Khi-Deux :

Si $\sum_{i=0}^{i=9} (fi - pi)^2 > \frac{2}{n}$ alors on rejette l'hypothèse d'adéquation, à la loi uniforme, des fréquences fi observées dans l'échantillon, avec un risque de 5%.
(ce test est applicable pour n assez grand, en pratique n > 30 et $n \times pi > 5$).

***

Dans l'expérimentation qui suit, le chiffre aléatoire exhibé est la première décimale du nombre aléatoire fourni par la fonction ALEA() du tableur (ou la fonction RAND de la calculatrice).

1. Comment obtenir la première décimale d'un nombre aléatoire fourni par la fonction ALEA() du tableur (ou la fonction RAND de la calculatrice) sachant que la fonction ALEA() (ou RAND) envoie un nombre appartenant à [0 ; 1[?
2. Générer un échantillon de taille n = 100
3. Calculer les fréquences f_0 , f_1 , ... , f_9.
4. Le test est il applicable pour n = 100 ? Si oui faire ce test
5. Recommencer avec n = 500 si la machine est équipée d'un programme automatisant les calculs de fréquences, ce qui est le cas sur tableur

COMMENTAIRES :
ce test justifie que la première décimale d'un nombre aléatoire de [0;1[, fourni par la fonction ALEA() de EXCEL ou par la fonction RAND de la calculatrice est un chiffre aléatoire; à contrario, le test ne justifie pas que l'on puisse utiliser l'échantillon de taille n comme table de chiffres aléatoires, voir pour cela le sujet n°3 .

COMPETENCES TICE :
* Utilisation de la fonction RAND où ALEA()
* Manipulation des listes

COMPETENCES MATHEMATIQUES
* La fonction partie entière
* Test d'adéquation à une loi uniforme

SUJET N°2

TEST D'ÉQUIRÉPARTITION SUR [0;1] D'UNE SUITE (u_n) .

Cet exercice peut se traiter sur tableur ou sur calculatrice programmable

soit $(u_n)_{n \in IN^*}$ une suite de nombres réels
[a;b] est un intervalle de IR
pour $n \in IN^*$, on note q(n) le nombre d'éléments de $\{u_1 , u_2 , ... , u_n\}$ appartenant à [a;b]
et f(n) = q(n)/n leur fréquence sur l'intervalle [a;b].
Définition : la suite $(u_n)_{n \in IN^*}$ est équirépartie sur [0;1] si, quel que soit [a;b] inclus dans [0;1], la suite $(f(n))_{n \in IN^*}$ tend vers b-a

****

1. $(u_n)_{n \in IN^*}$ est la suite définie par son terme général u_n = ALEA() (ou RAND sur la calculatrice), c'est à dire : u_1 = ALEA(), u_2 = ALEA(), ………… u_n = ALEA(),….
   On veut tester le caractère équiréparti de la suite $(u_n)_{n \in IN^*}$
   a. Mettre en mémoire les valeurs a = 1/3 et b = 1/2
   b. Créer une liste $\{u_1 , u_2 , ... , u_{100}\}$
   c. Programmer la création d'une liste L_1 de dimension 100, dans laquelle pour p = 1 ;2 ;….. ;100 $L_1(p) = 1$ si $u_p \in [a;b]$, et $L_1(p) = 0$ sinon, $L_1(p)$ désignant le terme de rang p de la liste L_1 .
   Demander au programme de renvoyer la somme de L_1 divisée par 100 qui est la fréquence cherchée (en d'autres termes, $\left(\sum_{p=1}^{p=100} L_1(p) \right) / \ 100$
   d. Pour [a;b] = [1/3;1/2], la définition donnée plus haut semble t'elle vérifiée par la suite $(u_n)_{n \in IN^*}$?
2. réitérer la procédure définie en 1. pour [a;b] = [1/2;1]
3. On sait que la fonction ALEA() (ou rand) envoie un nombre appartenant à [0 ; 1[.
   On rappelle que la partie entière d'un nombre réel x est l'unique entier n tel que $x \in [n ;n + 1[$. n = ENT(x) sur EXEL et n = Int(x) sur les calculettes TEXAS si elles sont en anglais.
   Pour tout intervalle [a;b] contenu dans [0 ; 1[on définit la probabilité p de l'événement « $u_n \in [a;b]$ » par: p = b-a. Expliquer alors pourquoi la fonction ENT(10 × ALEA()) (ou Int(10 × rand) renvoie un chiffre 0 ;1 ;2 ;… ;9 avec la probabilité 1/10 ?
4 . De même expliquer pourquoi, k étant un entier supérieur ou égal à 1, la fonction ENT(10^k × ALEA()) renvoie un nombre entier de k chiffres avec la probabilité $1/10^k$?
5. X_1 désigne la variable aléatoire égale à la première décimale de RAND ou ALEA() et X_2 la variable aléatoire égale à la deuxième décimale.
   a. Déduire de 4. les probabilités :
      - $p(X_1 = i)$ pour i = 0; 1 ;….. ;9
      - $p(X_1 = i$ et $X_2 = j)$, pour $i \in \{0; 1 ;….. ; 9\}$ et $j \in \{0; 1 ;….. ; 9\}$
   b. Calculer $p(X_2 = j)$, pour $j \in \{0; 1 ;….. ; 9\}$
   c. Les variables aléatoires X_1 et X_2 sont elles indépendantes ?

COMPETENCES TICE :
* Programmation
* Listes
* Fonction ALEA() ou RAND
COMPETENCES MATHEMATIQUE :
* Partie entière
* SUITES
* Probabilité
* Formule de probabilité totale et variables aléatoires indépendantes

SUJET N°3

TEST DU CARACTERE ALEATOIRE D'UNE TABLE DE CHIFFRES ALEATOIRES

L'exercice peut se traiter sur tableur ou sur calculatrice

On suppose que les suites obtenues en prenant successivement les 5 premières décimales des nombres donnés par :

- RAND ou ALEA() (constituant les termes de la suite (u_n))
- ou par $\sqrt{n}$ - E($\sqrt{n}$), où E(x) désigne la partie entière de x (constituant les termes de la suite (v_n)) ,

sont de bons candidats pour générer des suites de chiffres aléatoires .

Pour tester le caractère aléatoire d'une table (ou suite finie) de chiffres aléatoires, on regroupe ces chiffres par blocs de 5 qui sont répartis selon les 7 types décrits dans le tableau ci-dessous.

Dans ce tableau, les probabilités de chaque type ont été calculées, le test consiste alors à comparer la fréquence de chaque type dans la table avec la probabilité du même type donnée dans le tableau.

n°	description	type	exemple	probabilité
1	chiffres différents	abcde	34961	0,3024
2	une paire	aabcd	29512	0,5040
3	deux paires	aabbc	44533	0,1080
4	un triplet	aaabc	60366	0,0720
5	paire triplet	aaabb	23223	0,0090
6	quadruplet	aaaab	29222	0,0045
7	quintuplet	aaaaa	55555	0,0001

Ce test s'appelle le test du poker.

1. a. Combien de listes distinctes (de 5 chiffres chacune) peut-on constituer avec les chiffres 0, 1, 2, … , 9 , un même chiffre pouvant être répété jusqu'à 5 fois ?
   b. Quelle est la probabilité de chaque liste, le choix d'une liste se faisant « au hasard » ?
   c. Vérifier que ces listes peuvent être regroupées selon les 7 types décrits dans le tableau.
   d. Justifier, par le calcul, les probabilités données dans le tableau, le choix d'une liste se faisant « au hasard » ?

****

Dans l'expérimentation qui suit, on génère 100 blocs de 5 chiffres en prenant les 5 premières décimales données par RAND ou ALEA(), (donc la partie entière de $10^5 \times$ RAND ou de $10^5 \times$ ALEA()), puis on compare les fréquences observées de chacun des types aux probabilités théoriques données dans le tableau.

2. Générer une liste de 100 blocs de 5 chiffres obtenus comme décrit plus haut, avec RAND ou ALEA().

3. Calculer la fréquence de chacun des 7 types définis dans le tableau.

4. Comparer les fréquences observées aux probabilités théoriques données dans le tableau.
   La suite (u_n) semble t'elle passer le test du poker avec succès ?

5 Recommencer les étapes 2. 3. 4. avec la suite (v_n).

6. La suite (v_n) semble t'elle passer le test du poker avec succès ?

COMPETENCES TICE :
* Fonction RAND ou ALEA()
* Listes

COMPETENCES MATHEMATIQUES
* Calculs de probabilités

SUJET N°2

TEST D'ÉQUIRÉPARTITION SUR [0;1] D'UNE SUITE (u_n) .(variante graphique)

L'exercice peut se traiter sur tableur ou sur calculatrice programmable

soit $(u_n)_{n \in IN^*}$ une suite de nombres réels
[a;b] est un intervalle de IR
Pour $n \in IN^*$, on note q(n) le nombre d'éléments de $\{u_1, u_2, ..., u_n\}$ appartenant à [a;b]
et f(n) = q(n)/n , leur fréquence sur l'intervalle [a;b]

Définition : La suite $(u_n)_{n \in IN^*}$ est équirépartie sur [0;1] si, quel que soit [a;b] inclus dans [0;1], la suite $(f(n))_{n \in IN^*}$ tend vers b-a

---

Dans l'expérimentation qui suit, $(u_n)_{n \in IN^*}$ est la suite définie par son terme général u_n = ALEA()
(ou RAND sur la calculatrice),c'est à dire : u_1 = RAND , u_2 = RAND , u_3 = RAND , …, u_n = RAND,…..
On veut tester le caractère équiréparti de la suite $(u_n)_{n \in IN^*}$.

1. a. Mettre en mémoire les valeurs a = 1/3 et b = 1/2

   b. Créer une liste $\{u_1, u_2, ..., u_{100}\}$

   c. Créer une liste $L_1 = \{i_1, i_2, ..., i_p, ..., i_{100}\}$, où $\begin{cases} i_p = 1 \text{ si } u_p \in [a,b] \\ i_p = 0 \text{ sinon} \end{cases}$, p = 1 ; 2 ;…… ; 100.

   d. Créer une liste $L_2 = \{s_1, s_2, …, s_p, …, s_{100}\}$, où $s_p = \sum_{k=1}^{k=p} i_k$, p = 1 ; 2 ;….; 100.

   e. Créer une liste $L_3 = \{f(1), f(2), ..., f(p), ..., f(100)\}$, où $f(p) = \frac{s_p}{p}$, p = 1 ; 2 ;…… ; 100
   renvoyant la suite des fréquences $(f(p))_{1 \leq p \leq 100}$ des termes de $\{u_1, u_2, ..., u_p\}$ appartenant à [a;b].

   f. Utiliser les fonctions graphiques pour représenter sur [1; 100] les fonctions qui à p associe f(p) et à p associe b-a sur un même graphique.

   g. Pour [a;b] = [1/3;1/2], la définition donnée plus haut semble t'elle vérifiée par la suite $(u_n)_{n \in IN^*}$?

2. Réitérer la procédure définie au §1. pour [a;b] = [1/2;1].

---

3. On sait que la fonction ALEA() (ou RAND) envoie un nombre appartenant à [0 ; 1[.
   On rappelle que la partie entière d'un nombre réel x est l'unique entier n tel que
   $x \in [n ; n + 1[$, n = ENT(x) sur EXEL et n = Int(x) sur les calculettes TEXAS si elles sont en anglais.
   Pour tout intervalle [a;b] contenu dans [0 ; 1[on définit la probabilité p de l'événement
   « $u_n \in [a;b]$ » par: p = b-a. Expliquer alors pourquoi la fonction ENT(10 × ALEA())
   (ou Int(10 × RAND) envoie un chiffre 0 ;1 ;2 ;… ;9 avec la probabilité 1/10 ?

4. De même expliquer pourquoi, k étant un entier supérieur ou égal à 1, la fonction
   ENT(10^k × ALEA()) (ou Int(10^k × RAND) envoie un nombre entier de k chiffres avec la probabilité $1/10^k$?

---

5. X_1 désigne la variable aléatoire égale à la première décimale de RAND ou ALEA() et X_2 la variable aléatoire égale à la deuxième décimale.

a. Déduire de 4. les probabilités :
- $p(X_1 = i)$ pour $i = 0; 1; \ldots; 9$
- $p(X_1 = i \text{ et } X_2 = j)$, pour $i \in \{0; 1; \ldots; 9\}$ et $j \in \{0; 1; \ldots; 9\}$

b. Calculer $p(X_2 = j)$, pour $j \in \{0; 1; \ldots; 9\}$

c. Les variables aléatoires X_1 et X_2 sont elles indépendantes ?

COMPETENCES TICE :
* Programmation
* Listes
* Fonction ALEA() ou RAND
* Graphiques

COMPETENCES MATHEMATIQUES :
* Partie entière
* SUITES
* Probabilité
* Formule de probabilité totale et variables aléatoires indépendantes