

## Théorème central limite et hasard.

### Un script

```
a=floor(rand(1,16000)*2);
b=floor(rand(1,16000)*2);
c=floor(rand(1,16000)*2);
d=floor(rand(1,16000)*2);
r=8*a+4*b+2*c+d;
for i=0:15 do,
    sum(r==i),
end;
```

Examinons ce code d'un point de vue syntaxe du langage Scilab.

Il y a 5 variables, a à d et r. Ces variables ne sont pas déclarées. On peut supposer que ce sont des tableaux à 1 dimension et de longueur 16000. Soit au total 80000 valeurs stockées. On ne sait pas si ces valeurs sont sur 32 bits ou 64 bits, ou plus.

La fonction rand. La documentation précise que le résultat de l'appel est un nombre dans l'intervalle [0;1]. On sait que la méthode de calcul d'une valeur aléatoire est basée sur un calcul de reste de division d'entier, c'est "modulo". Pour obtenir un nombre dans l'intervalle [0;1], on divise le résultat du calcul par RANDMAX. On se trouve donc dans le cas où le nombre de valeurs possibles entre 0 et 1, considéré comme généralement infini est exactement égal à RANDMAX.

Après une exécution du script sur ma machine, on obtient les résultats suivants

```
983. 1010. 1039. 986. 1015. 1011. 939. 979.
968. 1033. 1053. 927. 1073. 1020. 972. 992.
```

Le calcul de la moyenne arithmétique, de l'écart\_type et de la répartition de ces écarts donne un résultat assez conforme à la répartition normale.

```
Nombre de valeurs = 16  valeur minimale =927.00  valeur maximale=1073.00
Rapport Emq/Ema = 1.21  Théorique = 1.25
Nombre = 16  Moyenne = 1000.00  emq=39.53  ep=26.35
```

```
Classe 1  nb=    0  0.00%  théorique 0.35% |
Classe 2  nb=    0  0.00%  théorique 2%   |
Classe 3  nb=    2  12.50% théorique 7%   | HHHHHHHHHHHHHHHH
Classe 4  nb=    2  12.50% théorique 16%  | HHHHHHHHHHHHHHHH
Classe 5  nb=    4  25.00% théorique 25%  | HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Classe 6  nb=    4  25.00% théorique 25%  | HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Classe 7  nb=    2  12.50% théorique 16%  | HHHHHHHHHHHHHHHH
Classe 8  nb=    2  12.50% théorique 7%   | HHHHHHHHHHHHHHHH
Classe 9  nb=    0  0.00%  théorique 2%   |
Classe 10 nb=    0  0.00%  théorique 0.35% |
```

### Programme C

```
int main()
{
    randomize();
    FILE *ecr=espion;
    int Res[16];
    for (int i=0; i<16; i++) Res[i]=0;
```

```

for (int i=0; i<16000; i++)
{
  int a=rand()%2;
  int b=rand()%2;
  int c=rand()%2;
  int d=rand()%2;
  int r=a*8 + b*4 + c*2 + d;
  Res[r]++;
}
for (int i=0; i<16; i++) fprintf(espion," %d",Res[i]);
return 0;
}

```

La logique est strictement la même, le tableau nécessaire n'utilise que 16 positions. Après exécution, on obtient le résultat suivant :

1051 973 1014 1033 1012 1020 1015 1002 995 983 976 987 976 986 1022 955  
Le calcul de moyenne, écart-type et répartition donne le résultats suivant :

Nombre de valeurs = 16 valeur minimale =955.00 valeur maximale=1051.00  
Rapport Emq/Ema = 1.17 Théorique = 1.25  
Nombre = 16 Moyenne = 1000.00 emq=25.47 ep=16.98

Classe	nb	pourcent	théorique
Classe 1	0	0.00%	théorique 0.35%
Classe 2	0	0.00%	théorique 2%
Classe 3	1	6.25%	théorique 7%
Classe 4	4	25.00%	théorique 16%
Classe 5	3	18.75%	théorique 25%
Classe 6	4	25.00%	théorique 25%
Classe 7	3	18.75%	théorique 16%
Classe 8	0	0.00%	théorique 7%
Classe 9	1	6.25%	théorique 2%
Classe 10	0	0.00%	théorique 0.35%

La comparaison de ces deux calculs permet les observations suivantes :

- 1- Dans les deux cas, la loi de probabilité utilisée est une loi uniforme. Les nombres calculés sont construits à partir d'un tirage type Pile ou Face.
- 2- Dans les deux cas, la moyenne des nombres de valeurs sorties est 1000, ce qui était prévisible. Ce qui vérifie la loi des grands nombres : sur un grand nombre de tirages de même loi, la moyenne tend vers la probabilité. Ce que l'on peut énoncer aussi de la façon suivante : "lors d'un grand nombre de mesures d'une même quantité, la moyenne arithmétique est la valeur la plus probable".
- 3- La symétrie "parfaite" de la répartition des résultats obtenus avec Scilab laisse planer un doute sur la fiabilité du résultat. En effet, l'écart-type est grand 39.53 alors qu'une valeur d'environ 25. est plus réaliste.
- 4- On vérifie dans les deux cas qu'une expérience réalisée avec une loi uniforme produit une résultat dont les écarts à la moyenne respectent la répartition dite normale.

Cette expérience de tirage aléatoire a été faite sous différentes formes un très grand nombre de fois. Les résultats ont toujours été conformes aux résultats théoriques des probabilités.

## Quelle loi de hasard ?

Cette question est souvent posée, cela sous-entendrait que la "loi de probabilité" devrait être connue pour tirer des conclusions et particulier, calculer la probabilité de tel événement. Cf. le paradoxe de Bertrand.

Pour soutenir cette affirmation il a été proposé l'énoncé suivant :

*Justement, chacun sa méthode, et cela influence les probabilités. Voici deux méthodes (parmi une infinité) pour choisir un nombre entre 1 et 100.*

*1ère méthode : on lance un dé à 100 faces (numérotée de 1 à 100) et on lit face présentée. Dans ce cas, vu que c'est un dé (supposé non pipé), on va dire que tous les nombres ont la même probabilité de sortir (à savoir 1/100) : c'est la loi uniforme sur les entiers de 1 à 100.*

*2ème méthode : une calculette permet d'obtenir un nombre X à virgule appartenant à l'intervalle ouvert ]0,1[ (suivant la loi uniforme). Ensuite on retient l'entier N compris entre 1 et 100 le plus proche de 1/X (le plus petit si 1/X est un nombre ...,5 ). Cela donne bien au "hasard" un entier N compris 1 et 100, mais cette manière de procéder ne donne pas lieu au même "hasard" qu'avec la première méthode, car les entiers n'ont pas les mêmes probabilités de sortir.*

### Comparaison des résultats

Avec la première méthode, on obtient ceci :

Nombre de valeurs = 100 valeur minimale =73.00 valeur maximale=128.00  
 Rapport Emq/Ema = 1.26 Théorique = 1.25  
 la valeur 128.000000 rang 6 est douteuse  
 Nombre = 100 Moyenne = 100.00 emq=10.21 ep=6.81

Classe 1	nb=	0	0.00%	théorique	0.35%	
Classe 2	nb=	1	1.00%	théorique	2%	H
Classe 3	nb=	7	7.00%	théorique	7%	HHHHHHH
Classe 4	nb=	19	19.00%	théorique	16%	HHHHHHHHHHHHHHHHHHHHH
Classe 5	nb=	24	24.00%	théorique	25%	HHHHHHHHHHHHHHHHHHHHH
Classe 6	nb=	21	21.00%	théorique	25%	HHHHHHHHHHHHHHHHHHHHH
Classe 7	nb=	18	18.00%	théorique	16%	HHHHHHHHHHHHHHHHHHH
Classe 8	nb=	7	7.00%	théorique	7%	HHHHHHH
Classe 9	nb=	2	2.00%	théorique	2%	HH
Classe 10	nb=	1	1.00%	théorique	0.35%	H

Nombre de valeurs = 100 valeur minimale =76.00 valeur maximale=127.00  
 Rapport Emq/Ema = 1.21 Théorique = 1.25  
 la valeur 127.000000 rang 27 est douteuse  
 Nombre = 100 Moyenne = 100.00 emq=9.79 ep=6.53

Classe 1	nb=	0	0.00%	théorique	0.35%	
Classe 2	nb=	3	3.00%	théorique	2%	HHH
Classe 3	nb=	6	6.00%	théorique	7%	HHHHHH
Classe 4	nb=	16	16.00%	théorique	16%	HHHHHHHHHHHHHHHHH
Classe 5	nb=	23	23.00%	théorique	25%	HHHHHHHHHHHHHHHHHHH
Classe 6	nb=	25	25.00%	théorique	25%	HHHHHHHHHHHHHHHHHHH
Classe 7	nb=	19	19.00%	théorique	16%	HHHHHHHHHHHHHHHHH
Classe 8	nb=	6	6.00%	théorique	7%	HHHHHH
Classe 9	nb=	1	1.00%	théorique	2%	H
Classe 10	nb=	1	1.00%	théorique	0.35%	H

Dans les deux cas, on obtient une répartition conforme à la répartition normale.

Le premier calcul a été fait en prenant au hasard un nombre entre 1 et 100.

Le second calcul a été fait en prenant au hasard un nombre entre 0 et 1, puis en comptant le nombre de sorties de chaque centile.

En réalité, dans l'énoncé, il y a la phrase "Ensuite on retient l'entier N compris entre 1 et 100 le plus proche de  $1/X$ ". Dès que l'on fait un tri des résultats fournis par le hasard, on intervient sur le hasard, donc, on n'est plus dans le cadre des probabilités qui sous-entendent par définition le caractère aléatoire d'une expérience.

Cependant, on peut avancer dans le raisonnement.

On réalise une expérience telle que décrite strictement dans la 2<sup>ème</sup> méthode. La courbe des résultats a une forme hyperbolique. On peut en écrire la fonction qui sera de la forme  $y = A / x$ . C'est ce qu'on appelle la loi de probabilité.

Maintenant, on réalise un grand nombre de fois cette expérience, en respectant la même loi, alors la répartition des écarts à la moyenne du paramètre A sera conforme à la répartition normale, puisque les résultats de toutes ces expériences ne dépendent que du hasard.

Un application classique de cela est l'étude de la durée de vie. On sait que ce phénomène a pour loi de probabilité une fonction exponentielle. L'utilisation est simple : " telle chose (une ampoule par exemple) a une chance sur deux de durer plus longtemps que la durée correspondant à la médiane, et non pas la moyenne ".

Si on intervient, directement ou indirectement, sur le résultat du hasard, l'expérience ne dépend plus seulement du hasard.

#### Autre exemple.

Voici un énoncé lu sur un forum.

Deux joueurs A et B lancent chacun leur tour un dé à 100 faces numérotées 1 à 100. Le joueur A a un handicap, s'il obtient plus de 50, son score est ramené à 50. Le vainqueur est celui qui obtient le meilleur score, en cas d'égalité, B gagne le coup. Quelle est la probabilité que le joueur A gagne ?

La "loi du hasard" est parfaitement définie, mais on remarque que l'issue d'un tirage est différente suivant que le joueur s'appelle A ou qu'il s'appelle B. On se trouve donc dans un cas comparable à celui de la corde de Bertrand où l'énoncé ne contient aucune ambiguïté, mais ici, l'exposé est suffisamment précis pour qu'il n'y ait pas d'interprétations différentes possibles. Par ailleurs, cet énoncé est tel que l'issue dépend du nom du joueur. Ceci étant précisé, on se trouve dans le cas caractérisé par l'expression "même loi".

Un calcul numérique, basé sur l'analyse combinatoire, donne une probabilité de gain pour le joueur A égale à 36.75%. On se trouve effectivement dans le cas peu fréquent où le résultat du calcul est un nombre "exact".

Il a paru intéressant de faire une simulation pour vérifier cela. C'est une application directe de la méthode de Monte-Carlo. En pareil cas, il y a deux façons de procéder, soit on fait un très grand nombre de tirages et on lit le résultat, soit on fait un certain nombre de fois un grand nombre de tirages et on examine les résultats. Il paraît évident que c'est la seconde méthode qu'il faut préférer, puisqu'on peut évaluer la dispersion et en déduire la qualité du résultat.

On considère que l'expérience réalisée est une partie de 10000 jeux. Il est utile, mais pas indispensable de choisir 10000 comme grand nombre, puisqu'on souhaite 4 chiffres significatifs.

Nombre de valeurs = 100 valeur minimale =3539.00 valeur maximale=3793.00  
 Rapport Emq/Ema = 1.21 Théorique = 1.25  
 Nombre = 100 Moyenne = 3676.23 emq=53.06 ep=35.37

Classe 1	nb= 0	0.00%	théorique 0.35%	
Classe 2	nb= 1	1.00%	théorique 2%	H
Classe 3	nb= 6	6.00%	théorique 7%	HHHHHH
Classe 4	nb= 18	18.00%	théorique 16%	HHHHHHHHHHHHHHHHHHH
Classe 5	nb= 29	29.00%	théorique 25%	HHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Classe 6	nb= 21	21.00%	théorique 25%	HHHHHHHHHHHHHHHHHHHHH
Classe 7	nb= 12	12.00%	théorique 16%	HHHHHHHHHH
Classe 8	nb= 10	10.00%	théorique 7%	HHHHHHHH
Classe 9	nb= 3	3.00%	théorique 2%	HHH
Classe 10	nb= 0	0.00%	théorique 0.35%	

### Interprétation des résultats.

Le rapport Emq/Ema est satisfaisant, ce qui vérifie la normalité du résultat. En d'autres termes, le générateur de nombres pseudo-aléatoires est correct et il n'y a pas de faute dans la simulation.

La moyenne = 36.7623% est tout à fait satisfaisante en comparaison de la probabilité calculée. Il s'agit là d'une vérification de la loi des grands nombres : "la moyenne tend vers la probabilité".

La valeur de l'emq (écart-type) = 0.53% paraît très importante pour un néophyte, en effet, cela signifie que le résultat est compris entre 36.23% et 37.29% pour un taux de confiance de 66%. Le valeur de l'écart-probable (ep=35.37 = 2/3 emq) correspond à une autre unité de calcul.

La répartition des 10 classes est une vérification du TCL : la répartition du résultat d'une expérience de même loi tend vers la loi normale.

Ci-dessous, la même expérience fait avec 1000 jeux au lieu de 10000 comme précédemment.

Nombre de valeurs = 100 valeur minimale =328.00 valeur maximale=401.00  
 Rapport Emq/Ema = 1.27 Théorique = 1.25  
 Nombre = 100 Moyenne = 365.81 emq=15.37 ep=10.24

Classe 1	nb= 0	0.00%	théorique 0.35%	
Classe 2	nb= 2	2.00%	théorique 2%	HH
Classe 3	nb= 9	9.00%	théorique 7%	HHHHHHHH
Classe 4	nb= 14	14.00%	théorique 16%	HHHHHHHHHHHH
Classe 5	nb= 24	24.00%	théorique 25%	HHHHHHHHHHHHHHHHHHHHHHHHHHH
Classe 6	nb= 26	26.00%	théorique 25%	HHHHHHHHHHHHHHHHHHHHHHHHHHH
Classe 7	nb= 16	16.00%	théorique 16%	HHHHHHHHHHHHHH
Classe 8	nb= 6	6.00%	théorique 7%	HHHHHH
Classe 9	nb= 3	3.00%	théorique 2%	HHH
Classe 10	nb= 0	0.00%	théorique 0.35%	

On observe que le gain sur la précision de la valeur moyenne est faible, ce qui justifie l'emploi de cette méthode connue sous le nom de Monte-Carlo.