TD n°1 de probabilités : de la fréquence expérimentale aux probabilités

CORRECTION

1. <u>Paradoxe du duc de Toscane</u>

Nous sommes au tout début de l'invention des probabilités : en 1620, à la cour de Florence, le duc de Toscane parie sur la somme de trois dés. Grand joueur, il a en effet remarqué que la somme 10 tombe un peu plus souvent que la somme 9, alors qu'on peut obtenir ces deux sommes du même nombre de façons différentes :

$9=6+2+1=5+3+1=5+2+2=4+4+1=4+3+2=3+3+3$ (six façons)

$10=6+3+1=6+2+2=5+4+1=5+3+2=4+4+2=4+3+3$ (six façons)

Il questionne Galilée (1554-1642) qui développe la question, jetant les bases de la théorie des probabilités.

a) L'approche statistique

Écrire un algorithme qui simule le tirage de trois dés à six faces, en calcule la somme S pour un nombre n de tirages et comptabilise les fois où S=9 et ceux où S=10.

>>Déterminer la fréquence expérimentale f_9 de l'événement « S=9 » et la fréquence f_{10} de l'événement « S=10 » pour un échantillon de $n=1000$ tirages. Comparer alors f_9 et f_{10}.

Cet algorithme en rappelle d'autres que nous avons fait au chapitre « Échantillonnage ».

Voici l'algorithme qui réalise l'objectif :

- N=0 (compte les tirages de somme Neuf), D=0 (compte les tirages de somme Dix)
- Pour J allant de 1 à 1000
  S=0 (enregistre la Somme des trois dés),
  Pour I allant de 1 à 3 : A=nombre aléatoire de [0;1[; S=S+*ent*(A×6+1)
  Si S=9 alors N=N+1
  Si S=10 alors D=D+1
- affichage de N/10 (fréquence des tirages de somme Neuf)
- affichage de D/10 (fréquence des tirages de somme Dix)

Remarques : nous utilisons un « nombre aléatoire de [0;1[» car c'est la forme la plus courante de nombre aléatoire (sauf erreur, cela se code *Ran#* sur Casio et *EntAleat* sur TI). Cela oblige, pour trouver un nombre entier de l'ensemble {1,2,3,4,5,6}, à utiliser l'expression *ent*(A×6+1) qui multiplie d'abord le « nombre aléatoire de [0;1[» pour se retrouver dans [0;6[; ensuite, on ajoute 1 pour être dans [1;7[; et enfin, on prend la partie entière du nombre pour obtenir ce qui est cherché.

Tout cela peut être avantageusement remplacé par la fonction appropriée connue par votre calculatrice (et par vous en principe), qui donne directement ce nombre entier. On remplacera alors :
A=nombre aléatoire de [0;1[; S=S+*ent*(A×6+1) par
A=nombre entier aléatoire entre 1 et 6 ; S=S+A

La fréquence est calculée par la formule N/10 et D/10 car il y a 1000 tirages et on multiplie par 100... Si on veut changer le nombre de tirages (noté T), on fera N×100/T et D×100/T.

En Python, et donc sur la Numworks, il faut importer le module *random* qui est désormais disponible (Adieu notre module bricolé avec sa problématique graine). Les fonctions de ce module qui sont utile : *random*() qui donne un pseudo réel aléatoire de [0;1[(ne pas oublier les parenthèses) et *randint*(1;6) qui donne un entier aléatoire (**rand**om **int**eger) de l'ensemble {1,2,3,4,5,6} (on peut bien sûr changer les bornes 1 et 6)

Programmons cet algorithme :

Casio	TI	Numworks

0→N
0→D
For 1→J TO 1000
0→S
For 1→I TO 3
Ran#→A
S+*ent*(A×6+1)→ S
Next
IF S=9
THEN N+1→ N
IfEnd
	0→N
	0→D
	For (J,1,1000)
	0→S
	For (I,1,3)
	EntAleat→A
	S+*ent*(A×6+1)→ S
	End
	IF S=9
	THEN N+1→ N
	End
		from random import *
		def troisdes(n) :
		    n9,n10=0,0
		    for i in range(n) :
		        s=0
		        for j in range(3) :
		            s+=randint(1,6)
		        if s==9: n9+=1
		        if s==10: n10+=1
		    print(« f9= »,n9/n,« f10= »,n10/n)

IF S=10 THEN D+1→ D IfEnd Next N/10 ◢ D/10 ◢	IF S=10 THEN D+1→ D End End Disp N/10 Disp D/10	

Exécutons ce programme : première remarque, c'est long.
Sur ma vieille Casio Graph25, il faut trois minutes pour les 1000 tirages.
J'obtiens 10,7 % pour la somme 9 et 12,9 % pour la somme 10.
Sur ma nouvelle Numworks c'est instantané (vive le progrès!). Je rappelle que pour exécuter cette méthode (def troisdes(n)) on se rend dans la console où on lance troisdes(1000). Pour éviter de retaper le nom, elle se trouve dans le menu *var* (le bouton var). J'obtiens l'affichage : f9= 0,124 f10=0,113 ce qui correspond à 12,4% pour f9 et 11,3% pour f10. Ce résultat étant un peu curieux (car il contredit ce qu'on a trouvé la 1ère fois et aussi l'observation du duc. Je relance en essayant n=10000. Cela prend quelques secondes troisdes(10000) me donne f9= 0,1103 f10=0,1291 ce qui correspond à 11,03% pour f9 et 12,91% pour f10.

Notez les noms de variables qui peuvent contenir plusieurs lettres en Python (et des chiffres). Dans le programme SIMULE vous pourrez ajouter d'autres simulations ; inutile d'avoir plusieurs programmes, un seul peut les contenir tous. La syntaxe d'une méthode : *def* nomdelamethode(variables) : suivi d'un retour à la ligne et d'une indentation (automatique en principe sur Numworks).

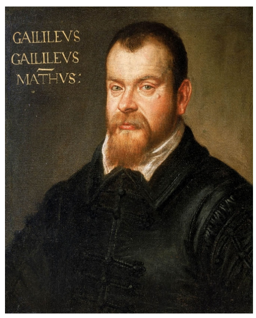
Ci-contre, voici une programmation en langage Python mais sur mon ordinateur. Dans ce dernier, j'ai ajouté une horloge qui mesure le temps d'exécution et j'ai profité de la facilité de saisie pour écrire des noms de variables plus explicite encore.
Vous pouvez constater que c'est plus rapide avec un ordinateur qu'avec une calculatrice : simuler 40 000 tirages ne prend même pas 1 seconde !

```
from random import randint
from time import clock
n=int(input('nombre de tirages = '))
temps=clock()
neuf,dix=0,0
for i in range(n):
    somme=0
    for j in range(3):
        dé=randint(1,6)
        somme+=dé
    if somme==9 :
        neuf+=1
    if somme==10 :
        dix+=1
print("réponse en {} secondes.".format(clock()-temps))
print("fréquence de la somme=9 :{}/{}={}.".format(neuf,n,100*neuf/n))
print("fréquence de la somme=10 :{}/{}={}.".format(dix,n,100*dix/n))
```

```
nombre de tirages = 40000
réponse en 0.39726917923065824 secondes.
fréquence de la somme=9 :4586/40000=11.465.
fréquence de la somme=10 :5135/40000=12.8375.
```

Avec 400 tirages, j'ai obtenu 9,5 % pour la somme 9 et 11,75 % pour la somme 10.
Avec 40000 tirages, j'ai trouvé 11,465 % pour la somme 9 et 12,8375 % pour la somme 10.

Ces essais confirment ce que le duc avait observé par l'expérience directe. Il devait avoir beaucoup de temps pour jouer (un bon sens de l'observation et une bonne mémoire) car avant d'observer avec certitude l'avantage du 10 sur le 9, il faut jouer plusieurs centaines de fois au minimum. Notre connaissance est aussi expérimentale, mais elle relève de la simulation (on ne tire pas de vrais dés) et utilise un outil dont le duc ne pouvait disposer : un ordinateur (ou une calculatrice) qui réalise tout cela (tirages et comptabilité) extrêmement vite.
Renseignement pris, le duc était grand duc et se prénommait Cosme, son titre étant Cosme II de Médicis. Il vivait à Florence entre 1590 et 1621 et fut instruit par son précepteur Galilée (1564-1642) entre 1605 à 1608. Le duc et Galilée sont restés de grands amis jusqu'à la mort prématurée du duc.



b) L'*intervalle de confiance* dans lequel on peut s'attendre à trouver la probabilité p au seuil de 5% est $[f_e-\frac{1}{\sqrt{n}}; f_e+\frac{1}{\sqrt{n}}]$ où f_e est la fréquence dans l'échantillon. Cet intervalle ayant une amplitude de $\frac{2}{\sqrt{n}}$, si l'on veut s'approcher de p à moins de 10^{-q}, il nous faut réaliser un nombre n d'expériences vérifiant $\frac{2}{\sqrt{n}}<10^{-q}$, soit $\sqrt{n}>2\times10^{q}$ ou encore $n>4\times10^{2q}$. Déterminer le nombre n de tirages à effectuer pour que la fréquence expérimentale soit une approximation de p au dixième près, puis au centième près.
Que pensez-vous alors de l'observation du duc ?
Si on veut avoir deux chiffres de précision, c'est-à-dire si on veut pouvoir distinguer des fréquences égales

à 0,11 et 0,12 par exemple, il faut prendre q=2 et $n>4\times 10^{2\times 2}=40000$. Il faut donc réaliser 40 000 tirages! Le duc est un grand joueur (il joue beaucoup) et/ou il a un don d'observation et une intuition hors du commun...
Sur la feuille j'avais écrit : si on veut que f_e approche p à moins de 10^{-q}, il faut réaliser un nombre n d'expériences tel que $\frac{1}{\sqrt{n}}<10^{-q}$.
En effet, si on donne f_e (le centre de l'intervalle) comme valeur approchée de p, on peut commettre une erreur d'au plus la moitié seulement de l'amplitude de l'intervalle. D'où une division par 2 pour $\sqrt{n}$ et par 4 de n.

c) L'arbre ci-dessous est difficile à tracer en entier mais il permet de faire la liste des chemins (branches) menant à la réalisation de ces événements.
>>Après avoir effectué ces dénombrements, déterminer les probabilités des événements « S=9 » et « S=10 ». Que pensez-vous qu'a répondu Galilée au duc ?
Il y a 6^3 possibilités équiprobables pour le tirage de 3 dés, soit 216 résultats ordonnés possibles.
Pour obtenir 9, on peut compter les permutations qui conviennent dans un arbre ou un tableau.
Faisons-en la liste :

621,**6**12,12**6**,1**6**2,2**6**1,21**6**	Le plus grand numéro est **6** : on ne peut faire 3 qu'avec 2 et 1
13**5**,1**5**3,31**5**,3**5**1,**5**13,**5**31 **5**22,2**5**2,22**5**	Le plus grand numéro est **5** : on peut faire 4 de deux façons, 2 et 2 ou 1 et 3
23**4**,2**4**3,32**4**,3**4**2,**4**23,**4**32 **4**41,**4**1**4**,1**44**	Le plus grand numéro est **4** : on peut faire 5 de deux façons, 2 et 3 ou 1 et 4
333	Le plus grand numéro est **3** : on ne peut faire 6 qu'avec 3 et 3

Cela fait 6+6+6+3+3+1=25 possibilités équiprobables.
Pour la somme égale à 10, on a 6+3+1, 5+4+1, 5+3+2 qui ont chacun 6 permutations et 6+2+2, 4+4+2 et 4+3+3 qui ont chacun 3 permutations. Cela fait en tout 6+6+6+3+3+3=27 possibilités équiprobables. Galilée a du répondre au duc qu'obtenir 10 est plus probable qu'obtenir 9. Sur 216 coups, en moyenne, il y en aura 2 de plus qui donneront une somme égale à 10 : pour être plus précis, 27 donneront la somme 10 et 25 la somme 9. Miser sur le 10 plutôt que sur le 9 assure, à long terme, une victoire assurée.
En termes modernes, on écrirait :
$P(S=9)=\frac{25}{216}\approx 0{,}11574$ alors que $P(S=10)=\frac{27}{216}=\frac{1}{8}=0{,}125$, d'où $P(S=10)>P(S=9)$.

2. Jeu équilibré ou jeu favorable/défavorable

Imaginons qu'à ce jeu, le joueur mise une somme m sur S_{10}, dans l'espoir de toucher k=6 fois la mise si les dés font 10. S'il perd, il perd sa mise. Par exemple, il mise m=5 € sur le 10 : si les dés font 10 il gagne 30 € (6×5=30) et récupère sa mise, dans tous les autres cas il perd sa mise de 5 €, donc il gagne −5 €.
a) Calculer ce que le joueur va gagner en moyenne, à chaque tirage des trois dés. Ce jeu est-il favorable ou défavorable au joueur ?
Pariant sur le 10, il gagnerait 27 fois sur 216 et gagnerait 30 €, le reste du temps il perdrait 5 €. Il gagnerait en moyenne $27\times 30+(216-27)\times(-5)=810-945=-135$. Le jeu serait donc défavorable au joueur.
Pariant sur le 9, il gagnerait 25 fois sur 216 et gagnerait 30 €, le reste du temps il perdrait 5 €. Il gagnerait en moyenne $25\times 30+(216-25)\times(-5)=750-955=-205$. Le jeu serait donc encore plus défavorable au joueur.

b) Pour qu'un jeu soit équilibré, il ne doit être ni favorable ni défavorable.
Calculer le facteur multiplicatif x pour que le jeu soit équilibré.
Pour que ce jeu soit équilibré en pariant sur le 10, il faudrait espérer gagner 0 €. Pour gagner 0 €, il faut gagner k fois la mise ; k doit vérifier $27\times k\times 5=(216-27)\times 5=945$ soit $k=\frac{945}{135}=7$. Il faut donc, pour équilibrer le jeu, proposer de gagner 7 fois la mise, donc 35 €. C'est facile à comprendre : il gagne une fois sur huit, comme il perd la mise 7 fois sur 8, il doit gagner 7 fois la mise lorsqu'il gagne.
Pour que ce jeu soit équilibré en pariant sur le 9, il faudrait espérer gagner 0 €. Pour gagner 0 €, il faut gagner k fois la mise ; k doit vérifier $25\times k\times 5=(216-25)\times 5=955$ soit $k=\frac{955}{125}=7{,}64$. Il faut donc, pour équilibrer le jeu, proposer de gagner 7,64 fois la mise, donc 38,2 €.