

TICE en Stats : « pourquoi pas PILE ou FACE ? »

Jean-Michel Puyou

IREM de Bordeaux

I Le problème initial

On lance 1000 fois de suite une pièce de monnaie non truquée.
Quelle est la probabilité d'obtenir au moins 548 Piles ?

Répondre sur un bout de papier.
Dépouillement.

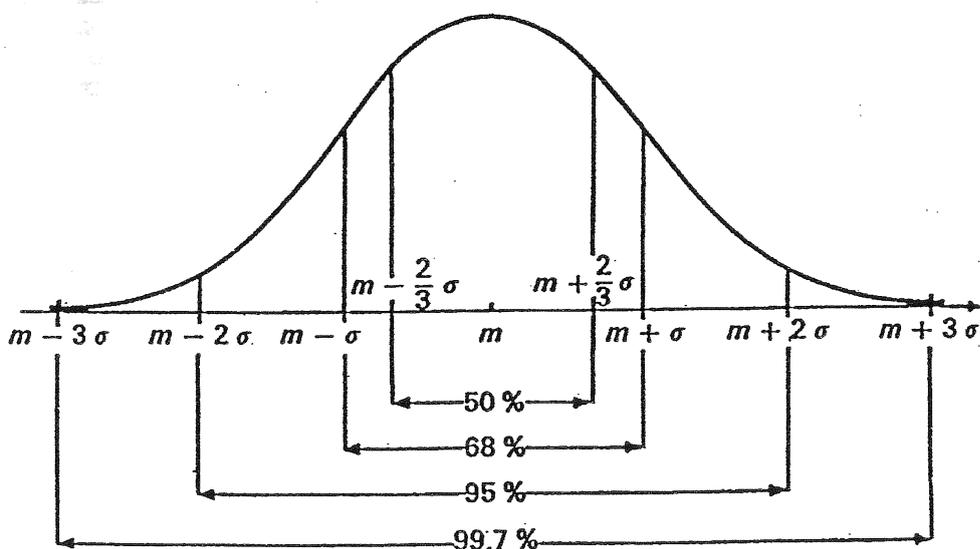
Solution de tête

On approche la loi binomiale $\beta\left(1000; \frac{1}{2}\right)$ par la loi normale $N(\mu, \sigma)$ avec $\mu = 500$

et $\sigma = \sqrt{1000 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{250} \approx 16$ (un peu moins).

Donc $548 \approx \mu + 3\sigma$ (légèrement supérieur).

La probabilité que le nombre de Piles soit à l'extérieur de $[\mu - 3\sigma ; \mu + 3\sigma]$ est de 3/1000 par raison de



symétrie la probabilité qu'il soit supérieur à $\mu + 3\sigma$ est de $1,5/1000$.

Par le calcul

Approximation par la loi Normale

Les conditions d'approximation de la loi binomiale par la loi normale sont satisfaites à savoir $n > 30$;
 $np > 5$; $nq > 5$

$T = \frac{X - np}{\sqrt{npq}}$ suit approximativement la loi normale $N(0 ; 1)$ ici $T = \frac{X - 500}{5\sqrt{10}}$.

Alors $X \geq 548 \Leftrightarrow T \geq \frac{48}{5\sqrt{10}} (\approx 3,03578655)$.

La fonction LOI.NORMAL.STANDARD d'EXCEL donne
 $1 - \text{NORM.S.DIST}(3,03578655), = 0,0012$

Utilisation directe de la loi binomiale

La fonction LOI.BINOMIALE cumulée d'EXCEL donne : 0.0013

II Situation en classe de seconde

1° Introduction

Que l'on ait 1 chance sur 2 d'obtenir Pile en lançant une pièce bien équilibrée n'est pas un scoop et se contenter de simuler des lancers de pièce pour atteindre ce résultat est un peu banal. Par contre le problème initial que je vous ai proposé réserve quelques surprises.

Nous avons transposé ce problème pour permettre de faire réaliser concrètement l'expérience par les élèves.

Voici l'énoncé proposé aux élèves de seconde :

On lance 100 fois de suite une pièce de monnaie non truquée.

Combien de chances a-t-on d'obtenir au moins 60 Piles ?

2° Déroulement d'une expérimentation concrète

➤ Transparent 1

La question posée aux 29 élèves de la classe, voici un tableau récapitulatif de leurs réponses

[0 ; 5 [2
[5 ; 10 [6
[10 ; 20 [8
[20 ; 30 [10
[30 ; 40 [0
[40 ; 50 [3

Je leur ai demandé de réaliser concrètement pendant les vacances de Pâques 2 séries de 100 lancers et de reporter les résultats dans une grille fournie.

Sur 54 séries réalisées on en dénombre 2 comportant 60 Piles ou plus.

On peut donc penser que la solution du problème est vraisemblablement dans l'intervalle [0 ; 5]

3° Simulation de 100 séries de 100 lancers sur EXCEL.(1^{er} niveau de simulation)

➤ Transparent 2

➤ Voir TD

Le recalcul de la feuille en utilisant la touche F9 fait apparaître une fluctuation de la fréquence de l'événement E entre 0 et 8 %.

Remarque : le nombre d'expérience a seulement doublé, par rapport à la manipulation des élèves et on n'a pas beaucoup gagné en précision.

4° Simulation de 1000 séries de 100 lancers

➤ Voir TD 4°)

a) pour chaque série on calcule la fréquence de Piles et on réalise un histogramme de ces fréquences. On constate une certaine symétrie du graphique et un centrage autour de 50%, ce qui est banal, l'allure de la courbe l'est moins.

➤ Transparent 3

b) Fourchette d'échantillonnage .

Illustration de la formule $P\left(\left|F_n - p\right| < \frac{1}{\sqrt{n}}\right) \geq 0,95.$

$F_n = \frac{S_n}{n}$ $T_n = \frac{S_n - np}{\sqrt{npq}}$ suit approximativement la loi normale $N(0 ; 1)$

Or $P(|T_n| \leq 1,96) \approx 0,95$.

Montrons que l'événement $(|T_n| \leq 1,96)$ est inclus dans $\left(|F_n - p| < \frac{1}{\sqrt{n}}\right)$.

En effet $|T_n| \leq 1,96 \Leftrightarrow \left|\frac{S_n - np}{\sqrt{npq}}\right| \leq 1,96 \Leftrightarrow |F_n - p| \leq \frac{1,96\sqrt{pq}}{\sqrt{n}}$.

Or avec $p + q = 1$, $pq = p(1 - q) \leq \frac{1}{4}$.

D'où $1,96\sqrt{pq} \leq 1,96 \times 0,5 \leq 1$.

Alors $|T_n| \leq 1,96 \Rightarrow |F_n - p| \leq \frac{1}{\sqrt{n}}$.

Ainsi $P\left(|F_n - p| < \frac{1}{\sqrt{n}}\right) \geq P(|T_n| \leq 1,96) \approx 0,95$.

> Voir TD échantillonnage

> Transparent 4

Pour nous $\left(|F_n - p| < \frac{1}{\sqrt{n}}\right) \geq 0,95 \Leftrightarrow P(0,4 < F_n < 0,6) \geq 0,95$.

La symétrie de l'histogramme des fréquences laisse prévoir que $P(F_n \geq 0,6) \approx 0,025$.

5°) Stabilisation des fréquences (2^{ième} niveau de simulation)

> Voir TD n° V

On crée une courbe représentant la fréquence d'apparition de l'événement E sur n série de 100 lancers en fonction de n.

> Transparent 5

On constate la stabilisation des fréquences autour d'une valeur proche de 0,028.

Le dessous des cartes est la loi faible des grands nombres :

Pour tout $\varepsilon > 0$ $P(|F_n - p| > \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$.

Problème : On fixe un tuyau autour de p [p - 0,01 ; p + 0,01] .

Comment choisir n pour que la probabilité que le point de coordonnées (n ; F_n) soit dans le tuyau .soit supérieure à 0,95 ?

Solution : On sait que $P\left(|F_n - p| < \frac{1}{\sqrt{n}}\right) \geq 0,95$.

Si $\frac{1}{\sqrt{n}} \leq 0,01$ on aura $P(|F_n - p| < 0,01) \geq P\left(|F_n - p| < \frac{1}{\sqrt{n}}\right) \geq 0,95$.

Il suffit donc de choisir $\sqrt{n} \geq 100$ c'est à dire $n \geq 10000$.

6°) Distribution des fréquences (3^{ième} niveau de simulation)

> Voir TD n° V

On récupère dans un tableau 100 fréquences de l'événement « obtenir 60 piles ou plus » obtenues par les élèves sur 1000 séries de 100 lancers.

On trace l'histogramme de la distribution de ces fréquences .

On obtient une courbe en cloche symétrique par rapport à la moyenne $\approx 0,28$.

➤ **Transparent 5**

Pour plus de précision on pourra à l'aide d'une macro-commande d'EXCEL, réaliser 1000 expériences de 1000 séries de 100.

```
Sub millefois()  
,
```

```
' millefois Macro
```

```
' Macro enregistrée le 01/06/04 par puyou  
,
```

```
Dim intcpt As Integer
```

```
For intcpt = 1 To 1000
```

```
Calculate
```

```
Cells(4 + intcpt, 106).Value = Cells(3, 103).Value
```

```
Next intcpt  
,
```

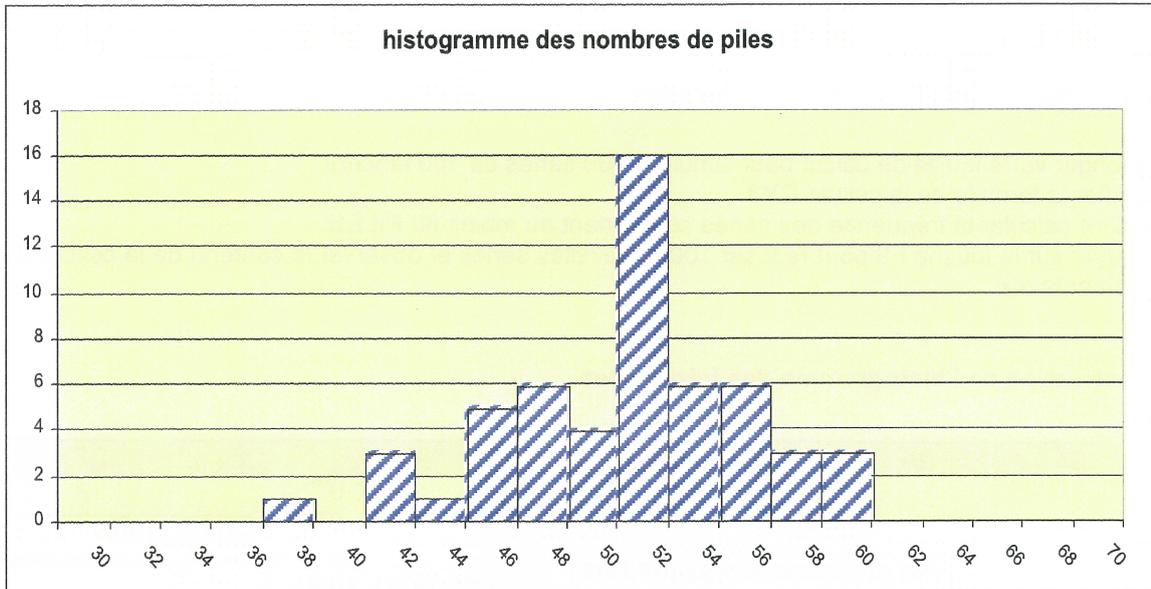
```
End Sub
```

TD Statistiques n°2

I La situation-problème

On lance 100 fois de suite une pièce bien équilibrée, quel pourcentage de chance a-t-on d'obtenir au moins 60 fois PILE ?

II Analyse des tirages effectués par les élèves



Sur 54 séries de 100 lancers réalisés concrètement 2 seulement atteignent 60 piles ou plus.

III Simulation sur Excel de 100 séries de 100 lancers

	A	B	C	D	E	F	G	H	CS	CT	CU	CV	CW	CX	CY	CZ
1																
2																
3															0,023	
4																
5	série	1	2	3	4	5	6	7	96	97	98	99	100	Nbre de P	>60	
6	1	0	0	0	0	0	0	0	1	1	1	0	1	49	0	0
7	2	1	0	0	0	0	0	1	0	0	1	1	0	55	0	0
8	3	1	0	0	0	0	1	0	0	0	1	0	0	53	0	0
9	4	0	1	0	0	1	0	0	1	0	0	1	0	48	0	0
10	5	0	1	1	0	0	0	0	0	1	0	1	1	52	0	0
11	6	1	0	0	0	1	1	1	1	0	1	1	1	63	1	1
12	7	1	1	0	0	1	1	1	1	1	0	1	0	44	0	1
97	92	0	1	0	0	0	0	0	1	0	0	0	1	44	0	5
98	93	1	1	1	1	1	1	0	1	0	1	0	1	51	0	5
99	94	1	1	1	0	1	0	0	1	1	0	1	1	51	0	5
100	95	0	1	1	1	0	1	1	1	0	0	0	1	50	0	5
101	96	0	1	0	0	0	0	1	0	1	1	1	1	48	0	5
102	97	0	0	0	0	0	0	0	1	1	0	1	0	47	0	5
103	98	1	0	0	0	1	1	1	0	0	0	1	1	50	0	5
104	99	1	0	0	0	0	1	0	0	0	0	1	0	45	0	5
105	100	1	0	0	1	0	0	1	0	1	1	0	1	55	0	5

De A6 à A105 inscrire les numéros des séries (penser à la recopie automatique)
 De B5 à CW5 inscrire les numéros des lancers de chaque série.
 En B6- entrer la formule = ENT(2*Alea())
 Recopier à droite jusqu'en CW6.
 En CX6 entrer la formule = somme(B6 :CW6)
 En CY6 entrer la formule = SI(CX6>=60 ;1 ;0)
 Sélectionner la plage B6 :CY6 et recopier jusqu'à la ligne 105.
 Ecrire en CY3 une formule permettant d'obtenir le nombre de séries au cours desquelles on a obtenu au moins 60 PILES.
 Appuyer sur la touche F9 pour réaliser 100 nouvelles séries et observer le contenu de la cellule CY3

IV Simulation de 1000 séries de 100 lancers

Prolonger votre feuille de calcul pour simuler 1000 séries de 100 lancers.
 Modifier la formule de la cellule CY3
 En CY4 calculer la fréquence des séries comportant au moins 60 PILES.
 Appuyer sur la touche F9 pour réaliser 1000 nouvelles séries et observer le contenu de la cellule CY4.
 Commentaires :

Construction de l'histogramme des fréquences

CZ	DA	DB	DC	DD	DE	DF
	30	<=30	0	0		
	31	<=31	0	0		
	32	<=32	0	0		
	33	<=33	0	0		
	34	<=34	0	0		
	35	<=35	2	0,002		
	36	<=36	3	0,001		
	37	<=37	3	0		
	38	<=38	8	0,005		
	39	<=39	20	0,012		
	40	<=40	28	0,008		
	63	<=63	999	0,005		
	64	<=64	1000	0,001		
	65	<=65	1000	0		
	66	<=66	1000	0		
	67	<=67	1000	0		
	68	<=68	1000	0		
	69	<=69	1000	0		
	70	<=70	1000	0		

Construire la colonne DA.
 En DB5 entrer le texte : « <=30 » et utiliser la recopie automatique pour compléter la colonne jusqu'en DB45 : « <=70 ».
 En colonne DC on calcule le nombre de séries correspondant au critère de la colonne DB.
 En colonne DD on calcule la fréquences du nombre de piles compris entre deux valeurs consécutives de la colonne DB.
 Représenter cet histogramme avec pour série des abscisses la colonne DA

Fourchettes d'échantillonnage.

Sur une nouvelle feuille recopier les 1000 séries de 100 lancers (colonnes A à CY)

	CS	CT	CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD
1												
2												
3						Fréquence	0,022					
4											Nbre de fourchettes contenant 0,5	960
5		96	97	98	99	100	Nbre de P >60	F100-0,1	F100+0,1	F100		
6	0	0	1	1	1		54	0	0,44	0,64	0,54	1
7	1	0	0	0	1		55	0	0,45	0,65	0,55	1
8	1	0	0	1	1		48	0	0,38	0,58	0,48	1
9	0	1	0	1	0		53	0	0,43	0,63	0,53	1
10	1	0	0	0	0		49	0	0,39	0,59	0,49	1
11	1	0	1	0	1		51	0	0,41	0,61	0,51	1
12	1	0	0	0	1		49	0	0,39	0,59	0,49	1
13	1	1	0	1	1		49	0	0,39	0,59	0,49	1
14	1	1	1	1	1		50	0	0,4	0,6	0,5	1
15	1	0	0	0	1		54	0	0,44	0,64	0,54	1
16	0	0	0	1	0		50	0	0,4	0,6	0,5	1
17	0	1	0	0	0		55	0	0,45	0,65	0,55	1
18	1	1	1	1	0		55	0	0,45	0,65	0,55	1
19	0	0	1	1	1		41	0	0,31	0,51	0,41	1
20	0	0	1	1	0		54	0	0,44	0,64	0,54	1
21	1	0	1	1	0		51	0	0,41	0,61	0,51	1
22	0	1	0	1	0		43	0	0,33	0,53	0,43	1
23	0	1	1	1	1		45	0	0,35	0,55	0,45	1
24	0	0	0	0	1		45	0	0,35	0,55	0,45	1
25	0	0	0	0	0		47	0	0,37	0,57	0,47	1
26	0	0	1	0	0		48	0	0,38	0,58	0,48	1
27	0	1	1	1	1		50	0	0,4	0,6	0,5	1
28	1	1	0	0	0		49	0	0,39	0,59	0,49	1

La colonne DB calcule la fréquence F100 de Pile de chaque série.

Les colonnes CZ et DA donnent pour chaque série les bornes de l'intervalle

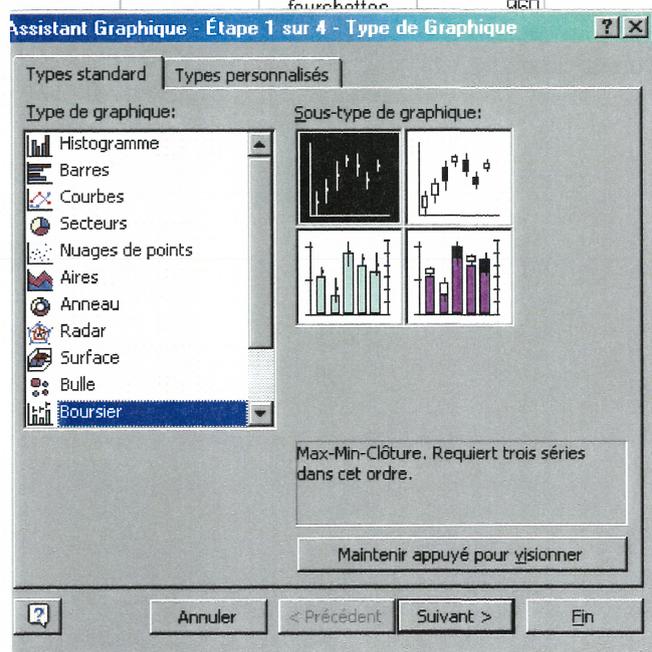
$$\left[F_{100} - \frac{1}{\sqrt{100}}; F_{100} + \frac{1}{\sqrt{100}} \right]$$

Représenter les colonnes CZ à DB dans un graphique boursier comme indiqué.

La théorie des statistiques nous dit que dans 95% des cas, cet intervalle contient la fréquence théorique 0,5.

Votre graphique semble-t-il vérifier cette théorie ?

Que peut-on en déduire pour la réponse à notre problème ?



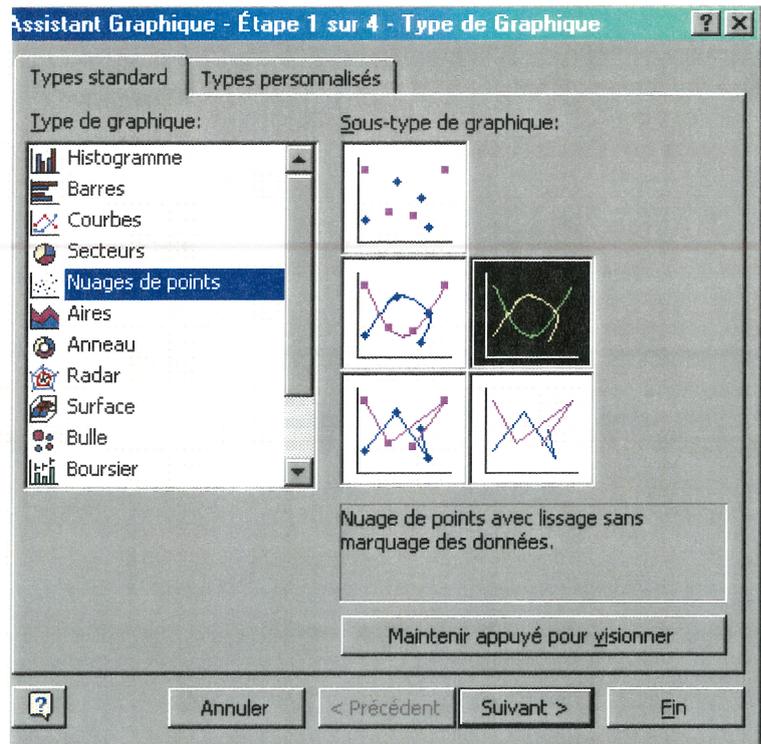
V Stabilisation des fréquences

Nous allons sur un nouvelle feuille réaliser 10000 séries de 100 lancers

	A	B	C	D	E	CU	CV	CW	CX	CY	CZ	DA
										0,0286		
série		1	2	3	4	98	99	100	Nbre de P	>=60	=CY6	=C26/A6
1	0	1	0	1	1	1	1		48	0	0	0
2	1	0	1	1	0	1	1		59	0	0	0
3	1	0	0	0	1	0	0		50	=C26+CY7	0	0
4	0	1	0	1	0	1	0		53	0	0	0
5	0	1	0	0	1	0	0		51	0	0	0
6	1	1	0	1	1	1	0		53	0	0	0
7	1	1	1	0	1	1	0		48	0	0	0
8	1	0	0	0	1	0	0		48	0	0	0
9	0	1	0	0	1	1	1		52	0	0	0
9995	0	1	1	1	0	0	0		48	0	285	0,02851426
9996	0	0	0	1	1	0	0		47	0	285	0,0285114
9997	1	0	0	1	1	1	0		53	0	285	0,02850855
9998	1	1	0	1	1	1	0		62	1	286	0,02860572
9999	0	1	1	1	0	1	1		56	0	286	0,02860286
10000	1	0	0	1	0	0	1		47	0	286	0,0286

La colonne CZ donne pour la série de rang n le nombre de fois que l'on obtient 60 piles ou plus au cours des n premières séries et la colonne DA donne la fréquence de cet événement sur les n premières séries.

Représenter sous forme de nuage de points les colonnes A et DA.
Que constate-t-on ?



VI Distribution des fréquences de l'événement « obtenir 60 piles ou plus »

Reprendre la feuille 1 et noter la fréquence de l'événement « obtenir 60 piles ou plus » calculée par la cellule **CY3**.

Appuyer sur la touche F9 pour recalculer la feuille de calcul noter la nouvelle fréquence.
Recommencer une fois encore cette opération.

Nous allons récupérer dans un tableau 100 fréquences ainsi obtenues.

	A	B	C	D	E	F	G	H
1								
2		Distribution des fréquences						
3								
4								
5	expérience: 1000 séries de 100	Fréquence de "Nb Piles >=60		Intervalles	critère	effectifs cumulés	effectifs	
6	1			0,015	<=0,015	0	0	=F6
7	2			0,017	<=0,017		0	=F7-F6
8	3			0,019	<=0,019			
9	4			0,021	<=0,021			
10	5			0,023	<=0,023			
11	6			0,025	<=0,025			
12	7			0,027	<=0,027			
13	8			0,029	<=0,029			
14	9			0,031	<=0,031			
15	10			0,033	<=0,033			
16	11			0,035	<=0,035			
17	12			0,037	<=0,037			
18	13			0,039	<=0,039			
19	14			0,041	<=0,041			
20	15			0,043	<=0,043			
21	16			0,045	<=0,045			
22	17							
99	94							
100	95							
101	96							
102	97							
103	98							
104	99							
105	100							
106								

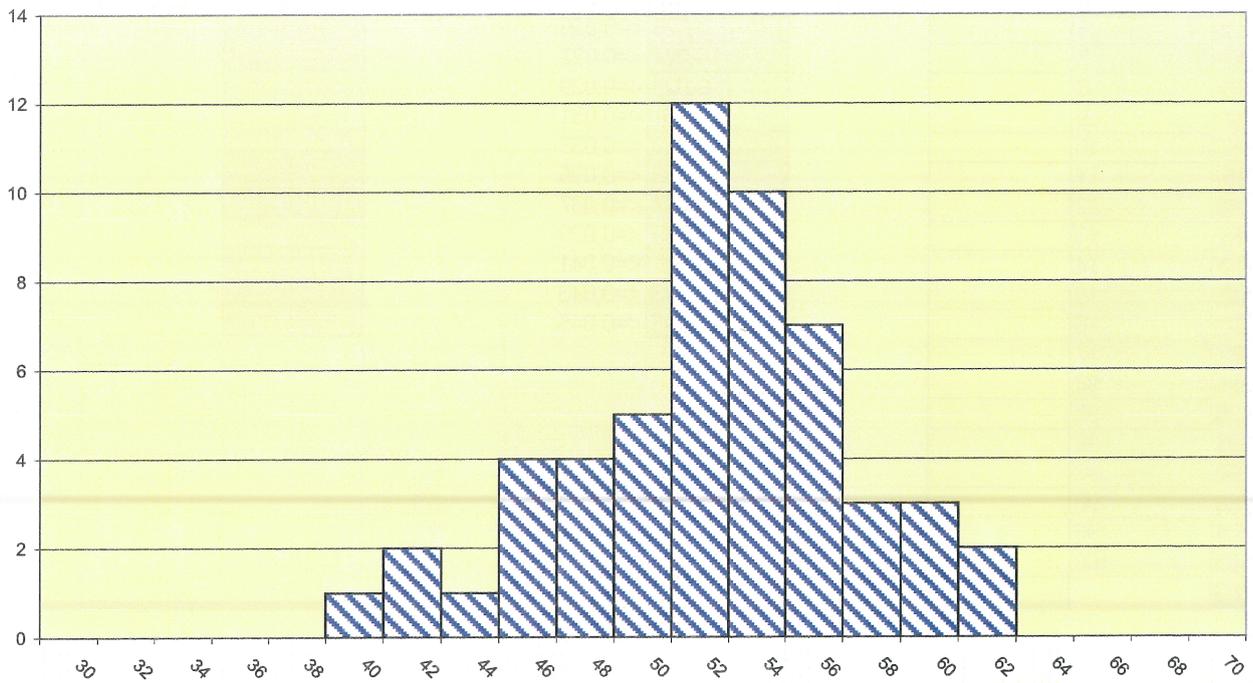
Représenter par un histogramme la colonne G en prenant pour série des abscisses la colonne D
Observer l'allure du graphique obtenu

Sondage dans une classe de seconde

[0 ; 5 [2
[5 ; 10 [6
[10 ; 20 [8
[20 ; 30 [10
[30 ; 40 [0
[40 ; 50 [3

Résultats des manipulations des élèves

Histogramme du nombre de piles

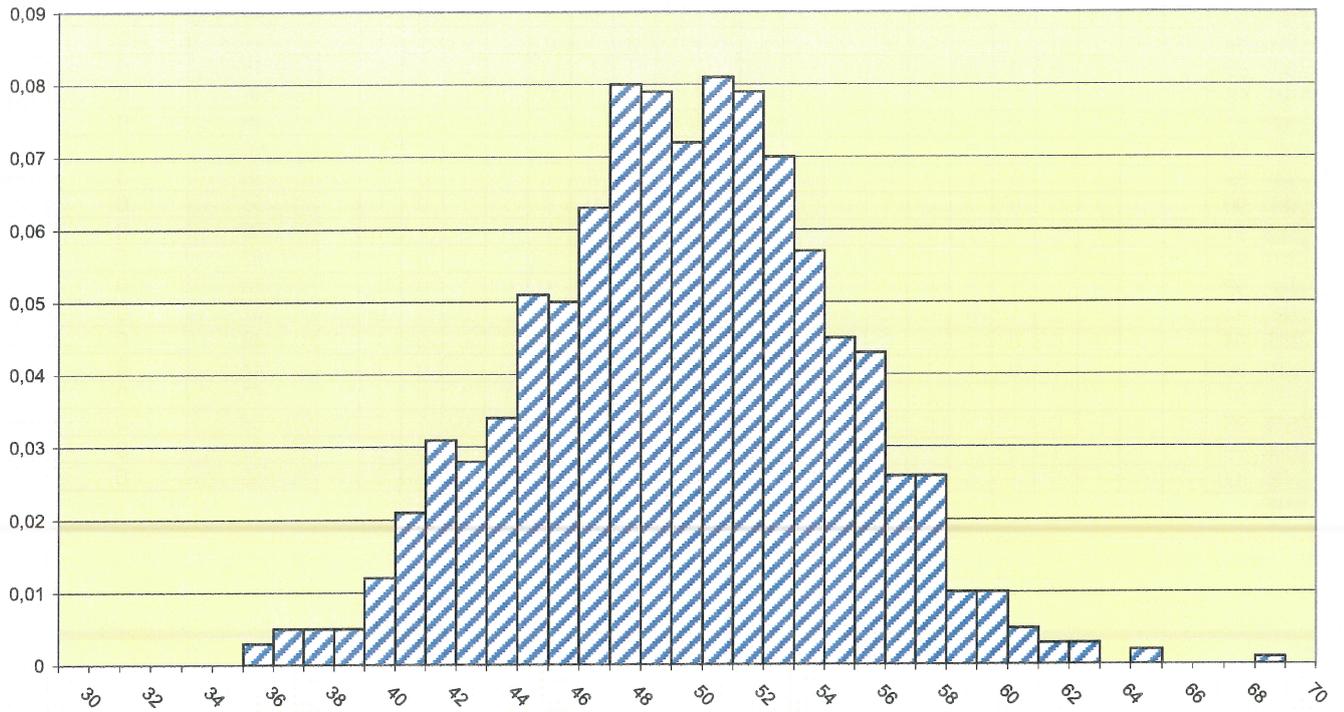


Feuille 1
100 séries de 100 lancers

	A	B	C	D	E	F	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW	CX	CY	
1																									
2																								= SOMME(\$CY\$6:\$CY\$105)/100	
3																									
4																									
5	série	1	2	3	4	5	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	Nbre de R	>60	
6	1	1	0	0	1	1	1	1	0	1	0	1	0	1	1	1	1	0	0	0	1	0			
7	2	1	1	1	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	1	0			
8	3	1	0	0	1	0	0	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1			
9	4	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	1	0	0	0	1	1			
10	5	0	1	0	1	1	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	1			
11	6	1	1	1	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1			
12	7	1	1	0	1	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	1			
13	8	0	1	0	1	1	0	0	0	1	1	1	0	0	0	0	0	1	0	0	1	0			
87	82	1	1	1	1	0	1	0	1	0	0	0	1	1	1	0	0	1	0	1	0	0			
88	83	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0			
89	84	0	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	1	0			
90	85	0	1	0	0	0	0	0	1	1	1	0	1	0	0	1	0	1	0	0	0	0			
91	86	0	0	0	1	1	1	0	1	0	1	0	1	1	0	1	0	0	0	1	1	0			
92	87	1	1	0	0	0	0	0	1	0	0	1	1	1	1	1	0	0	0	0	0	0			
93	88	1	0	0	1	1	0	1	0	1	1	0	0	0	0	1	1	1	1	1	0	1			
94	89	1	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0			
95	90	1	0	1	0	0	1	0	1	0	1	0	0	1	0	0	1	1	1	1	0	1			
96	91	0	1	1	0	1	1	1	1	1	0	0	0	1	1	1	1	0	0	1	1	1			
97	92	1	1	0	1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	0	1	1			
98	93	1	1	0	1	1	0	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0			
99	94	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0			
100	95	1	0	0	0	0	1	0	0	0	1	1	1	1	0	0	1	1	0	0	1	0			
101	96	1	0	1	1	0	1	1	1	0	1	0	1	0	0	0	1	0	0	0	1	0			
102	97	1	1	0	0	1	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	1			
103	98	0	1	0	0	1	1	0	0	1	0	1	0	1	0	1	0	0	0	1	1	1			
104	99	1	1	1	1	0	1	0	1	1	0	0	0	1	1	1	0	1	0	0	0	1			
105	100	0	0	1	0	1	0	0	0	1	1	1	1	1	1	0	0	1	0	0	0	0			

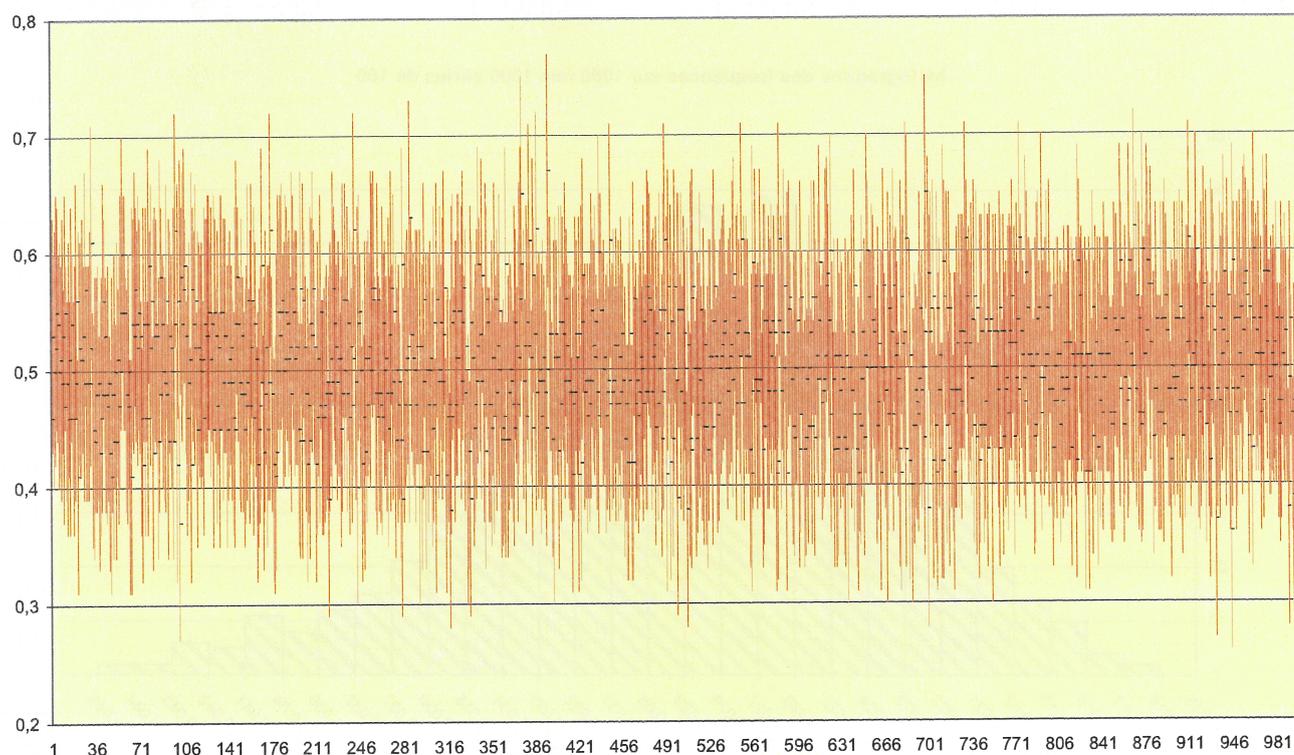
**Distribution des fréquences du nombre de pile
sur 1000 séries de 100 lancers**

histogramme des fréquences

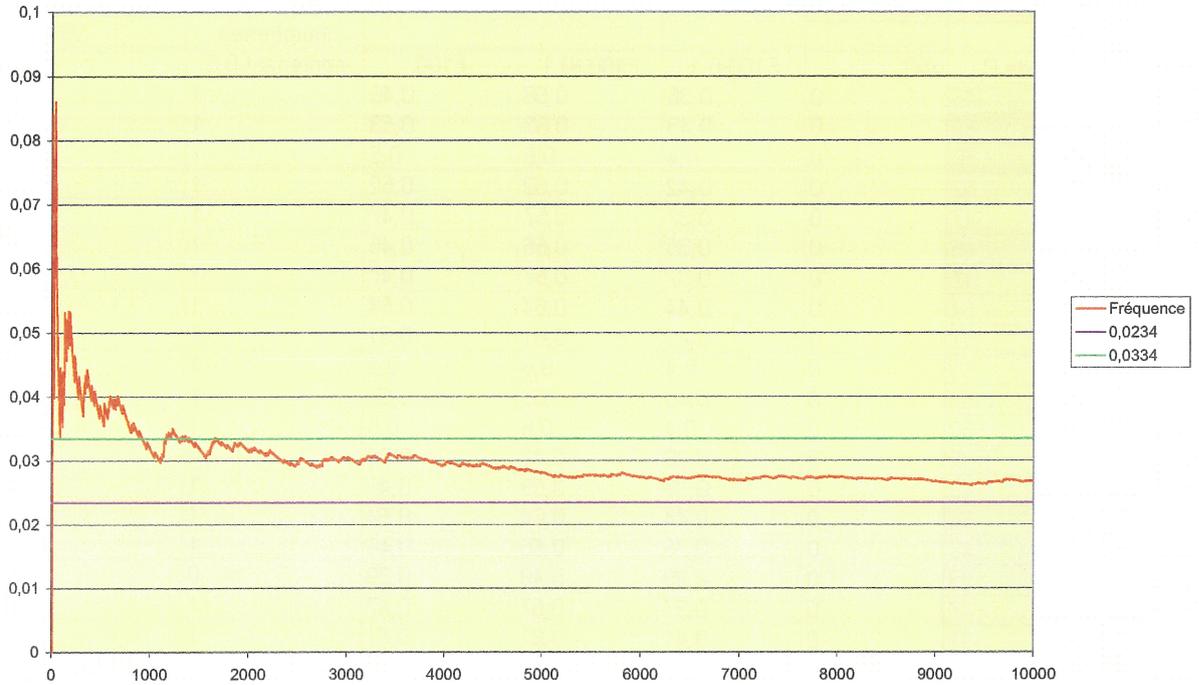


Fourchette d'échantillonnage au seuil de 95 %

CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD
			Fréquence	0,028				Nbre de fourchettes contenant 0,5	952
98	99	100	Nbre de P	>60	F100-0,1	F100+0,1	F100		
0	0	1	46	0	0,36	0,56	0,46	1	
1	1	0	53	0	0,43	0,63	0,53	1	
1	0	0	50	0	0,4	0,6	0,5	1	
1	1	1	52	0	0,42	0,62	0,52	1	
0	1	0	47	0	0,37	0,57	0,47	1	
0	0	0	45	0	0,35	0,55	0,45	1	
0	0	0	47	0	0,37	0,57	0,47	1	
1	0	1	54	0	0,44	0,64	0,54	1	
0	0	1	51	0	0,41	0,61	0,51	1	
1	0	1	50	0	0,4	0,6	0,5	1	
1	1	0	53	0	0,43	0,63	0,53	1	
0	1	1	50	0	0,4	0,6	0,5	1	
1	0	0	38	0	0,28	0,48	0,38	0	
1	0	0	49	0	0,39	0,59	0,49	1	
1	0	0	54	0	0,44	0,64	0,54	1	
1	1	1	46	0	0,36	0,56	0,46	1	
1	0	1	39	0	0,29	0,49	0,39	0	
1	1	1	47	0	0,37	0,57	0,47	1	
1	0	1	51	0	0,41	0,61	0,51	1	
0	1	1	55	0	0,45	0,65	0,55	1	
0	0	0	43	0	0,33	0,53	0,43	1	
1	1	0	56	0	0,46	0,66	0,56	1	



Stabilisation des fréquences d'apparition de l'événement E



1000 expériences de 1000 séries de 100 lancers

histogramme des fréquences sur 1000 fois 1000 séries de 100

