

# INTERVALLE DE FLUCTUATION AU SEUIL DE 95 % D'UNE LOI BINOMIALE

Auteur : Christian Vassard

TI-83 Premium CE

Fichiers associés : fluctuation\_loi binomiale\_eleve.pdf, IF.8xp, IF2.8xp

## 1. Objectifs

La situation est la suivante : une loi binomiale  $X$  de paramètres  $n$  et  $p$  étant donnée, déterminer un intervalle de fluctuation  $I$  à 95 % de cette variable aléatoire, c'est-à-dire un intervalle  $[a; b]$ , où  $a$  et  $b$  sont des entiers compris entre 0 et  $n$ , ayant une probabilité d'au moins 95 % d'être réalisé par la variable aléatoire.

On écrira un algorithme pour déterminer l'intervalle de fluctuation demandé dans les classes de premières générales de lycée. On cherchera comment améliorer cet algorithme.

On examinera quelques exemples d'application à la prise de décision.

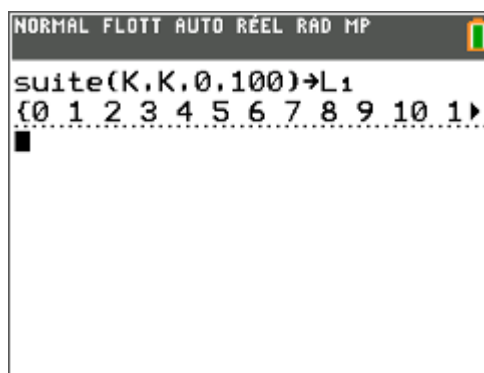
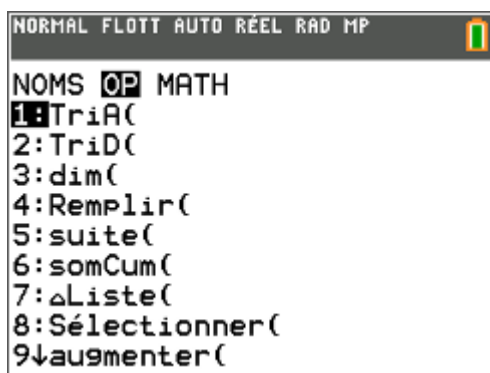
## 2. Plusieurs intervalles de fluctuation...

1)  $X$  est une variable aléatoire suivant la loi binomiale de paramètres  $n = 100$  et  $p = 0,3$ .

a) Comme la variable aléatoire  $X$  prend ses valeurs dans l'intervalle  $[0; n]$ , il est clair que  $P(0 \leq X \leq n) = 1 \geq 0,95$ , donc que l'intervalle  $[0; n]$  est un intervalle de fluctuation des valeurs de la loi binomiale au seuil de 95 %. Cet intervalle est de bien peu d'intérêt, on en convient.

b) Intéressons-nous à un intervalle de la forme  $[0; c]$  où  $c$  est un entier naturel compris entre 0 et  $n$ . Reprenons la méthode décrite dans l'énoncé.

On commence par mémoriser dans la liste  $[L1]$  la liste des entiers de 0 à 100, pour que la lecture des probabilités soit facilitée. L'instruction **suite** s'obtient dans le menu **[2nde] [stats]** et on choisit l'onglet **OP** :



Dans la liste  $L1$ , on met les valeurs cumulées de la loi binomiale, autrement dit pour  $k$  variant de 0 à 100, les valeurs  $P(X \leq k)$ . On va cette fois dans le menu **[2nde] [var]** (soit **[distrib]**) et on choisit **binomFRéP** : quand on ne choisit comme paramètres que  $n = 100$  et  $p = 0,3$ , sans préciser de valeur particulière, la calculatrice renvoie dans une liste l'ensemble des valeurs correspondant à la loi binomiale considérée.

Remarquons que pour cette fonction comme pour la précédente, on dispose d'une aide en ligne dès lors que dans le **[mode]**, les **ASSISTANTS STATS** sont réglés à **AFF**.

```
NORMAL FLOTT AUTO RÉEL RAD MP
DISTR DESSIN
4↑invT(
5:studentFdp(
6:studentFRép(
7:χ²Fdp(
8:χ²FRép(
9:FFdp(
0:FFRép(
A:binomFdp(
B:binomFRép(
```

```
NORMAL FLOTT AUTO RÉEL RAD MP
suite(K,K,0,100)→L1
{0 1 2 3 4 5 6 7 8 9 10 1}
binomFRép(100,0.3)→L2
{3.23447651E-16 1.4185489}
◀
```

Il reste à aller dans l'éditeur de listes (**stats**) pour examiner d'un peu plus près les résultats obtenus.

On constate alors que la plus petite valeur de  $c$  telle que l'intervalle  $[0 ; c]$  soit un intervalle de fluctuation au seuil de 95 % est  $c = 38$ .

On a :  $P(0 \leq X \leq 38) \approx 0,96602 \geq 0,95$ .

L1	L2	L3	L4	L5	1
30	.54912				
31	.63311				
32	.71072				
33	.77926				
34	.83714				
35	.88392				
36	.92012				
37	.94695				
38	.96602				
39	.97901				
40	.9875				

L1(39)= 38

c) Cet intervalle de fluctuation répond bien à la définition proposée mais il a le défaut d'inclure des valeurs dont la probabilité est faible ; on doit pouvoir trouver facilement un intervalle de fluctuation de longueur moins grande.

Décalons simplement cet intervalle d'une unité, en considérant l'intervalle  $[1 ; 39]$ . Cette fois :

$$P(1 \leq X \leq 39) = P(X \leq 39) - P(X \leq 0) \approx 0,9790.$$

```
NORMAL FLOTT AUTO RÉEL RAD MP
binomFRép(100,0.3,39)
.9790114234
binomFRép(100,0.3,0)
3.23447651E-16
.9790114234-3.23447651E-16
.9790114234
```

On peut raccourcir cet intervalle notablement, en choisissant par exemple de conserver la borne supérieure 39.

Le plus petit intervalle de fluctuation au seuil de 95 % que l'on trouve est alors :  $[22 ; 39]$ .

```
NORMAL FLOTT AUTO RÉEL RAD MP
binomFRép(100,0.3,39)
.9790114234
binomFRép(100,0.3,21)
.0288312528
.9790114234-.0288312528
.9501801706
```

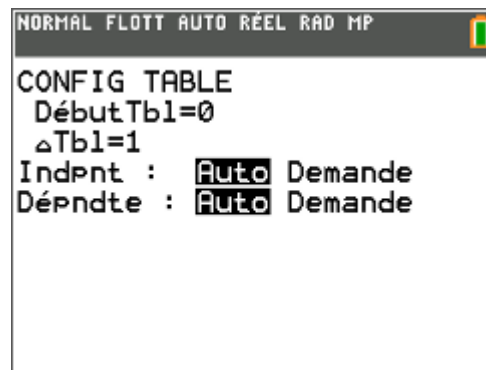
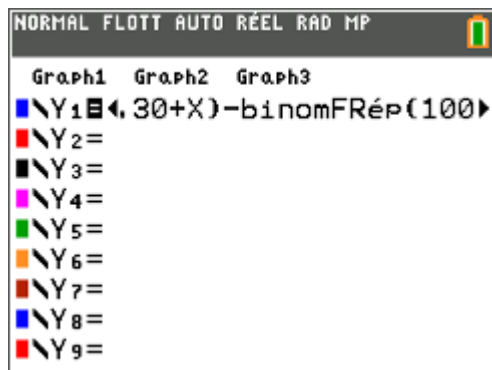
2) a) L'espérance mathématique de la variable aléatoire  $X$  est  $E(X) = np = 100 \times 0,30 = 30$ .

b) On cherche cette fois-ci un intervalle centré sur l'espérance 30, de la forme  $[30 - s; 30 + s]$ .

Comme le suggère l'énoncé, saisissons dans  $Y_1$  la fonction définie par :

$$\text{binomFRéP}(100,0.3,30+X) - \text{binomFRéP}(100,0.3,30-X-1)$$

et observons dans la table (après l'avoir configurée comme ci-dessous) le moment où cette fonction dépasse 0,95.



NORMAL FLOTT AUTO RÉEL RAD MP					
APP SUR + POUR ΔTbl					
X	Y1				
0	.08678				
1	.25633				
2	.41435				
3	.55486				
4	.67401				
5	.77035				
6	.84459				
7	.89909				
8	.93719				
9	.96255				
10	.97861				
X=0					

On constate que c'est pour  $X = 9$ , ce qui conduit à l'intervalle  $[30 - 9; 30 + 9] = [21; 39]$ .

Trois approches différentes, trois intervalles de fluctuations distincts : ont une légitimité plus grande les intervalles les plus courts, c'est-à-dire ceux qui mettent de côté les valeurs de probabilités les plus faibles. On conçoit cependant bien ce que la notion a précisément de *fluctuant* et qu'il faudra s'en tenir à une définition précise, celle du programme.

### 3. L'intervalle de fluctuation du programme

Le programme des classes de premières S et ES demande de considérer l'intervalle  $[a; b]$  tel que :

$a$  est le *plus petit entier* tel que  $P(X \leq a) > 2,5 \%$

$b$  est le *plus petit entier* tel que  $P(X \leq b) \geq 97,5 \%$

1) a) Tout d'abord, remarquons que :

$$P(X \in [a; b]) = P(a \leq X \leq b) = P(X \leq b) - P(X \leq a-1).$$

D'une part, on sait par hypothèse que :

$$P(X \leq b) \geq 97,5 \% \quad (1)$$

D'autre part,  $a$  est le *plus petit entier* tel que  $P(X \leq a) > 2,5 \%$  ; on est donc sûr que  $P(X \leq a-1) \leq 2,5 \%$ . De cette inégalité, on déduit que :

$$-P(X \leq a-1) \geq -2,5 \% \quad (2)$$

En ajoutant membre à membre les inégalités (1) et (2), il vient :

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a-1) \geq 0,975 - 0,025 = 0,95$$

ce qui montre que l'intervalle  $[a; b]$  est un intervalle de fluctuation au seuil de 95 % de la loi binomiale.

b) Il est clair que l'événement  $a \leq X \leq b$  est le même que l'événement  $\frac{a}{n} \leq \frac{X}{n} \leq \frac{b}{n}$ . On a donc aussi

$P\left(\frac{a}{n} \leq \frac{X}{n} \leq \frac{b}{n}\right) = P(a \leq X \leq b) \geq 0,95$ . L'intervalle de fluctuation de la fréquence du succès au seuil de 95 % est  $\left[\frac{a}{n}; \frac{b}{n}\right]$ .

2) a) Comme plus haut, on utilise les listes et **binomFrep**.

On obtient les écrans suivants :

```
NORMAL FLOTT AUTO RÉEL RAD MP
suite(K,K,0,100)→L1
{0 1 2 3 4 5 6 7 8 9 10 1}
binomFRep(100,0.3)→L2
{3.23447651E-16 1.4185489}
```

L1	L2	L3	L4	L5	1
15	4E-4				
16	9.7E-4				
17	.00216				
18	.00452				
19	.00889				
20	.01646				
21	.02883				
22	.04787				
23	.07553				
24	.11357				
25	.16313				

L1(22)= 21

L1	L2	L3	L4	L5	1
29	.46234				
30	.54912				
31	.63311				
32	.71072				
33	.77926				
34	.83714				
35	.88392				
36	.92012				
37	.94695				
38	.96602				
39	.97901				

L1(40)= 39

L'intervalle de fluctuation au seuil de 95 % de la loi binomiale est donc l'intervalle [21 ; 39]. La probabilité de cet intervalle dépasse 0,95. La loi binomiale prend essentiellement ses valeurs dans cet intervalle.

b) On procède de même pour l'autre exemple.

L1	L2	L3	L4	L5	1
501	.00456				
502	.00553				
503	.0067				
504	.00807				
505	.00969				
506	.01158				
507	.01378				
508	.01634				
509	.0193				
510	.0227				
511	.02659				

L1(512)= 511

L1	L2	L3	L4	L5	1
559	.90802				
560	.91882				
561	.92863				
562	.93751				
563	.94551				
564	.95268				
565	.95907				
566	.96475				
567	.96976				
568	.97417				
569	.97804				

L1(570)= 569

L'intervalle de fluctuation de la variable aléatoire  $X$  est l'intervalle [511 ; 569], correspondant à une fluctuation de la variable aléatoire fréquence de  $[0,511; 0,569]$ .

3) a) Le programme complété est le suivant :

```

Saisir n et p
K prend la valeur 0
R prend la valeur  $P(X \leq K)$ 
Tant que  $R \leq 0,025$ 
    K prend la valeur  $K+1$ 
    R prend la valeur  $P(X \leq K)$ 
Fin du tant que
A prend la valeur K
Afficher A
Tant que  $R < 0,975$ 
    K prend la valeur  $K+1$ .
    R prend la valeur  $P(X \leq K)$ 
Fin du tant que
B prend la valeur K
Afficher B
Afficher  $P(A \leq X \leq B)$ 

```

Ce programme reprend le principe de notre démarche : on balaie une par une toutes les valeurs de **K**, depuis **0** jusqu'à **N**, et on observe le franchissement d'abord du seuil 2,5 %, puis du seuil 97,5 %.

On remarque les inégalités qui définissent la condition de maintien dans la boucle : inégalité large dans la première, inégalité stricte dans la deuxième. En effet, dans le premier cas, on doit sortir de la boucle dès que  $P(X \leq K) > 0,025$  donc on y reste tant que  $P(X \leq K) \leq 0,025$  ; dans le second cas, on sort de la boucle dès que  $P(X \leq K) \geq 0,975$  donc on y reste tant que  $P(X \leq K) < 0,975$ .

b) Le programme est la suivant :

```

NORMAL FLOTT AUTO RÉEL RAD MP
PROGRAM: IF
:Prompt N,P
:0→K
:binomFRép(N,P,K)→R
:While R≤0.025
:K+1→K
:binomFRép(N,P,K)→R
:End
:K→A
:Disp A

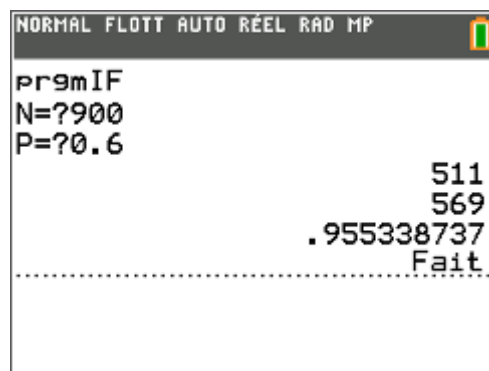
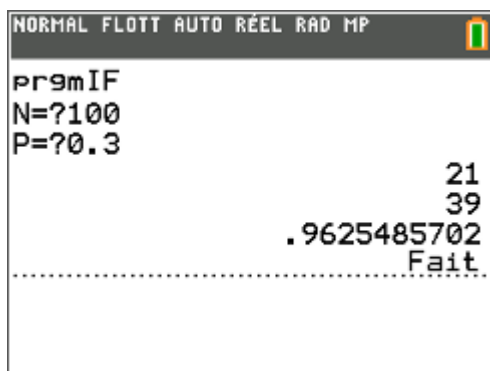
```

```

NORMAL FLOTT AUTO RÉEL RAD MP
PROGRAM: IF
:Disp A
:While R<0.975
:K+1→K
:binomFRép(N,P,K)→R
:End
:K→B
:Disp B
:Disp binomFRép(N,P,B)-binomFRép(N,P,A-1)

```

c) Les résultats obtenus sont les suivants :



On retrouve notre résultat pour le premier exemple. Le deuxième peut demander 2 minutes sur la calculatrice.

#### d) Prolongement facultatif

C'est une partie que l'on peut aborder avec des élèves motivés... ou agacés de la lenteur de leur calculatrice. Comment peut-on améliorer cet algorithme ?

Il suffirait simplement de ne pas balayer toutes les valeurs de l'intervalle  $[0; n]$  comme on le fait actuellement.

Deux remarques peuvent alors guider notre réflexion.

Tout d'abord, le fait que les probabilités les plus grandes pour une loi binomiale se rencontrent autour de l'espérance  $np$ .

Ensuite, on peut penser à l'intervalle de fluctuation vu en seconde. Rappelons-en les termes : il suppose que l'on répète  $n$  fois ( $n$  étant supérieur ou égal à 25) une épreuve au cours de laquelle un événement de probabilité  $p$  (comprise entre 0,2 et 0,8) peut être réalisé ou non. On sait qu'alors dans 95 % des cas, la fréquence de réalisation de l'événement est dans l'intervalle  $\left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$ .

On peut transposer tout ceci à une loi binomiale de paramètres  $n$  et  $p$ , pour laquelle  $n \geq 25$  et  $0,2 \leq p \leq 0,8$ .

La variable aléatoire fréquence  $\frac{X}{n}$  se trouve dans l'intervalle  $\left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$  avec une probabilité d'au moins 0,95. Autrement dit :

$$P\left(p - \frac{1}{\sqrt{n}} \leq \frac{X}{n} \leq p + \frac{1}{\sqrt{n}}\right) \geq 0,95$$

ce qui équivaut à :

$$P(np - \sqrt{n} \leq X \leq np + \sqrt{n}) \geq 0,95$$

Voilà donc un intervalle où se concentrent les probabilités les plus fortes de la loi binomiale : c'est lui qu'il faut viser, peu ou prou.

On peut donc suggérer de commencer l'exploration non pas à 0, comme on le fait dans l'algorithme précédent mais un peu avant la partie entière de  $np - \sqrt{n}$  ou à 0 si ce dernier nombre est négatif.

De même, dès qu'on a obtenu **A** on peut faire un saut un peu avant la partie entière de  $np + \sqrt{n}$ .

Cela donne l'algorithme suivant :

```
NORMAL FLOTT AUTO RÉEL RAD MP
PROGRAM: IF2
:Prompt N,P
:max(0,ent(N*P-1.3*J(N))→K

:binomFRép(N,P,K)→R
:Disp R
:While R≤0.025
:K+1→K:binomFRép(N,P,K)→R
:End
:K→A
```

```
NORMAL FLOTT AUTO RÉEL RAD MP
PROGRAM: IF2
:ent(N*P+0.7*J(N))→K
:binomFRép(N,P,K)→R
:Disp R
:While R<0.975
:K+1→K:binomFRép(N,P,K)→R
:End
:K→B
:Disp A,B
:Disp binomFRép(N,P,B)-bin
```

```
:Disp binomFRép(N,P,B)-bin
omFRép(N,P,A-1)
```

Comprenons-nous bien : nous ne disposons d'aucune garantie absolue d'obtenir le bon résultat, c'est la raison pour laquelle il est prudent de faire afficher les probabilités  $P(X \leq K)$  pour les deux valeurs que nous choisissons : la première doit être inférieure à 0,025, la seconde à 0,975. Sinon, il faut modifier les bornes du balayage... ou faire tourner l'ancien algorithme, malgré sa lenteur.

On obtient les résultats suivants en des temps très raisonnables, une quinzaine de secondes pour le second, une trentaine pour le troisième, à comparer avec l'ancienne performance pour le deuxième.

```
NORMAL FLOTT AUTO RÉEL RAD MP
PRgmITZ
N=?100
P=?0.3

.0021629332
.9469544148
21
39
.9625485702
Fait
```

```
NORMAL FLOTT AUTO RÉEL RAD MP
PRgmITZ
N=?900
P=?0.6

.0045552119
.9286303828
511
569
.955338737
Fait
```

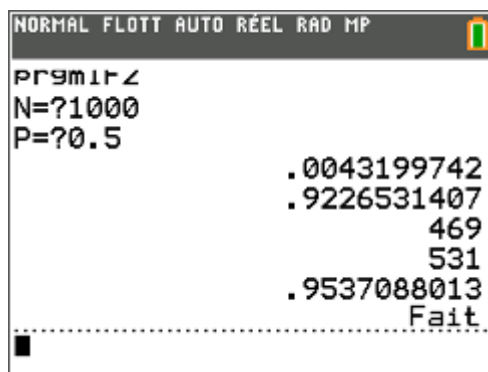
```
NORMAL FLOTT AUTO RÉEL RAD MP
PRgmITZ
N=?3000
P=?0.7

.0023227051
.9379246362
2051
2149
.9514302103
Fait
```

### 3. Application à une prise de décision

1) a) Le nombre de pile au cours des 1 000 lancers suit bien une loi binomiale de paramètres  $n = 1\,000$  et  $p = \frac{1}{2}$ , car on suppose la pièce parfaite.

On peut donner l'intervalle de fluctuation au seuil de 95 % à l'aide de la calculatrice :



Cet intervalle est [469 ; 531].

b) Si Théo prétend avoir obtenu 560 « pile » au cours des 1 000 lancers, tout dépend de l'hypothèse que l'on veut tester, qu'il est prudent de formuler avant.

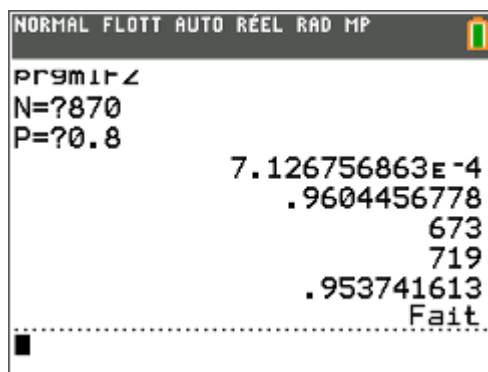
On peut soit prétendre que Théo est un menteur et qu'il n'a pas effectué vraiment les lancers, au risque de 5 % de nous tromper, soit déclarer que la pièce n'est pas parfaite, avec le même risque de se tromper.

2) L'exercice est classique et figure dans les documents d'accompagnement. Il était fastidieux de le traiter à la calculatrice mais le second algorithme écrit règle ces problèmes de lenteur.

On considère la loi binomiale  $X$  donnant le nombre d'Américains d'origine hispanique dans un jury d'un comté du Texas.  $X$  a pour paramètres  $n = 870$  et  $p = 0,8$ .

On fait l'hypothèse que la proportion d'Hispano-Américains est de 80 % dans un tel jury et on se demande si une telle répartition peut survenir.

Déterminons l'intervalle de fluctuation au seuil de 95 % de la variable aléatoire  $X$ .



C'est [673 ; 719]. Les 339 personnes de notre exemple donnent un nombre largement en dehors de notre intervalle de fluctuation au seuil de 95 %. On peut donc rejeter l'hypothèse d'une répartition des Hispano-Américains dans le jury, conforme à leur proportion dans la population.