Echantillonnage - Intervalle de fluctuation

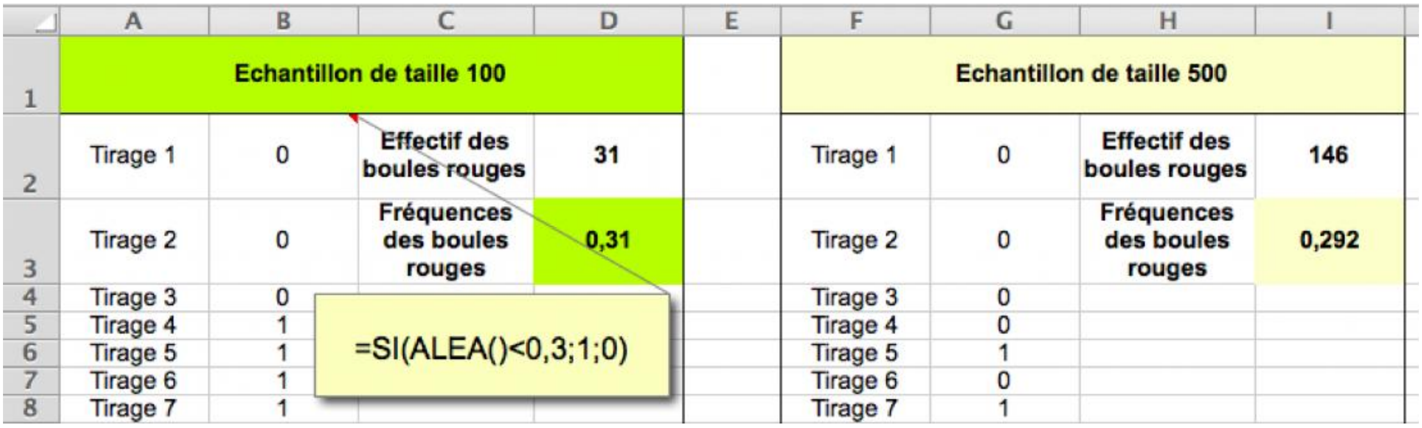
A. Rappel - Intervalle de fluctuation

Définition : Supposons que la proportion d'un caractère dans une population est connue et vaut p. Un intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n, est un intervalle dans lequel se situe la fréquence observée f dans un échantillon de taille n avec une probabilité environ égale à 0,95.

Propriété : Soit p la proportion effective d'un caractère d'une population et f la fréquence du caractère dans un échantillon de taille n. Si p est compris entre 0,2 et 0,8 et n est supérieur ou égale à 25 alors f appartient à l'intervalle $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$ avec une probabilité d'environ 0,95.

Expérimentation : Nous disposons d'un biberon opaque qui contient des boules rouges et des boules qui ne sont pas rouges. Le contenu de ce biberon peut être assimilé à une population dans laquelle on connait la proportion p de boules rouges qui est égale à 0,3. En retournant ce biberon, on peut observer la couleur de la boule sortante, ce qui permet de réaliser rapidement un tirage. En répétant cette manipulation n fois, on effectue n « tirages indépendants » et on obtient ainsi un échantillon de taille n. Le nombre de boules rouges obtenu (ou la fréquence des boules rouges) dans un échantillon de taille n est variable : c'est cette « fluctuation », que nous allons observer et analyser.

Simulation avec tableur : Pour obtenir des échantillons de grande taille, on va utiliser un tableur. Voici un extrait de la feuille de calcul à créer :



	A	B	C	D	E	F	G	H	I
1	Echantillon de taille 100					Echantillon de taille 500			
2	Tirage 1	0	Effectif des boules rouges	31		Tirage 1	0	Effectif des boules rouges	146
3	Tirage 2	0	Fréquences des boules rouges	0,31		Tirage 2	0	Fréquences des boules rouges	0,292
4	Tirage 3	0				Tirage 3	0		
5	Tirage 4	1				Tirage 4	0		
6	Tirage 5	1				Tirage 5	1		
7	Tirage 6	1				Tirage 6	0		
8	Tirage 7	1				Tirage 7	1		

=SI(ALEA()<0,3;1;0)

1.a) Expliquer la formule entrée dans les cellules B2 et G2. Jusqu'à quelle ligne faut-il recopier cette formule ?
b) Quelles formules faudra-t-il entrer dans les cellules D2, D3, I2 et I3 ?
2. En appuyant sur la touche F9, compléter les tableaux suivants, (page 2), avec 24 fréquences de boules rouges sur les échantillons de taille 100, puis de taille 500.
3. Sur le graphique 1, placer les points de coordonnées $(i\,;f_i)$ en rouge pour les échantillons de taille 100 et en bleu pour les échantillons de taille 500.
4. Placer sur le graphique 2 les points ayant pour coordonnées $(n\,;f_i)$.

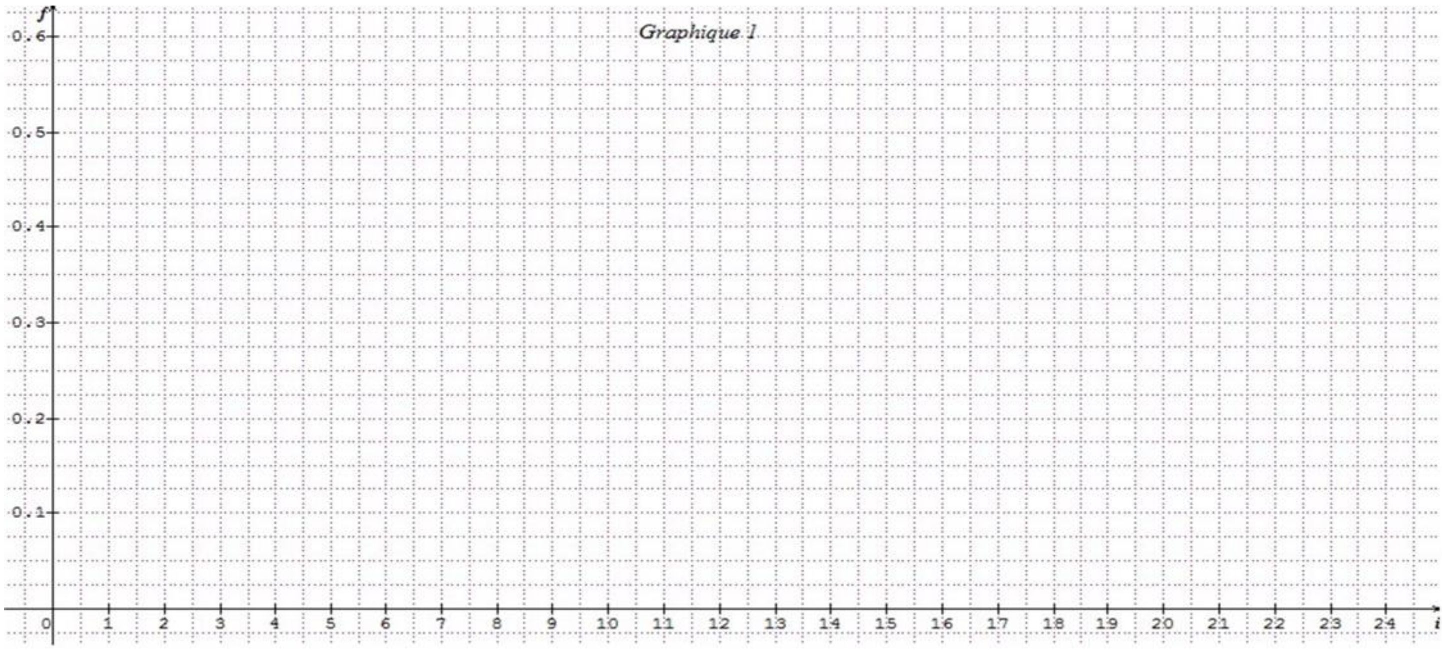
Interprétation :
1. Quel est l'effet de la taille de l'échantillon sur les résultats obtenus ?
2. Sur le graphique 2 :

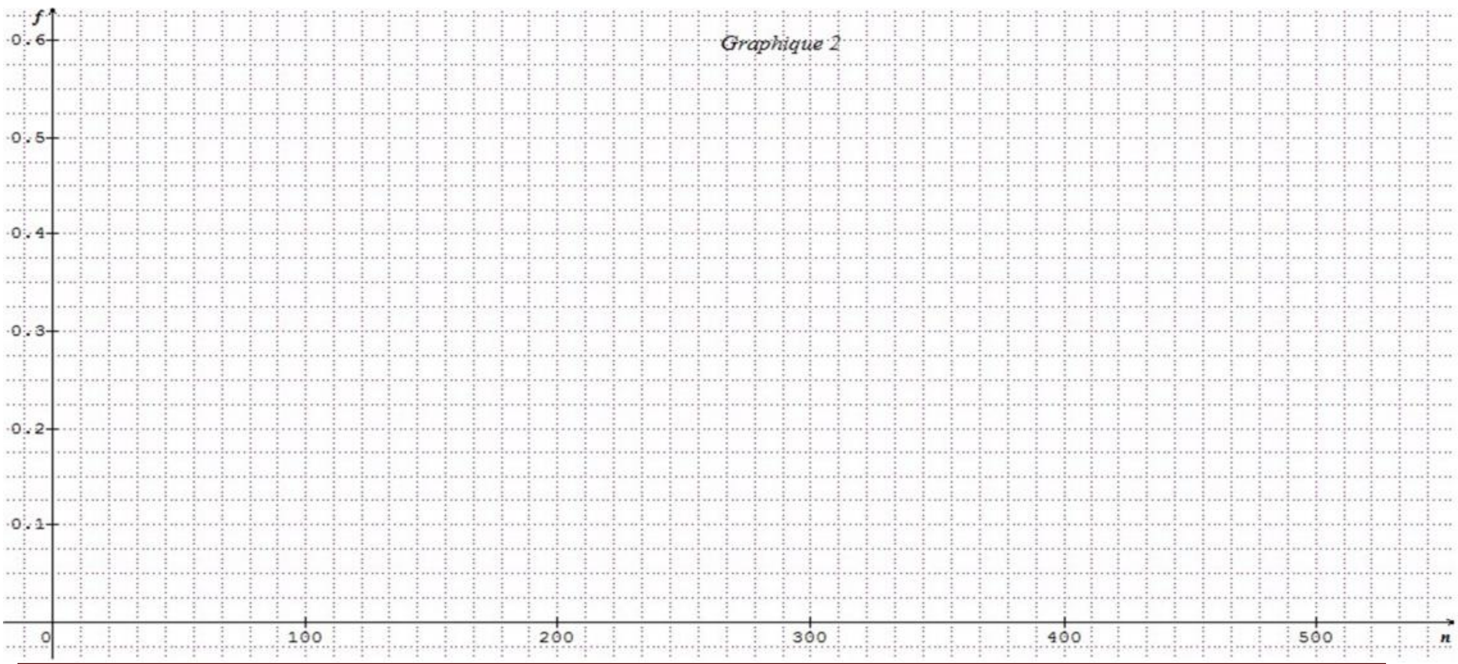
a) Placer les points A et B de coordonnées respectives $\left(n;0,3-\frac{1}{\sqrt{n}}\right)$ et $\left(n;0,3+\frac{1}{\sqrt{n}}\right)$ pour n = 100.

b) Idem pour les points C et D pour n = 500.
c) Déterminer le pourcentage de points appartenant aux segments [AB] et [CD]. Retrouver ces pourcentages en exploitant le graphique 1.

d) Conclure par rapport à la propriété donnée en début d’activité.

Taille n = 100												
Echantillon i	1	2	3	4	5	6	7	8	9	10	11	12
Fréquence f_i												
Echantillon i	13	14	15	16	17	18	19	20	21	22	23	24
Fréquence f_i												
Taille n = 500												
Echantillon i	1	2	3	4	5	6	7	8	9	10	11	12
Fréquence f_i												
Echantillon i	13	14	15	16	17	18	19	20	21	22	23	24
Fréquence f_i												





B. Une autre méthode

Définition : Soit X une variable aléatoire qui suit une loi binomiale $\mathcal{B}(n\,;p)$ et $F=\dfrac{X}{n}$ la variable aléatoire qui représente la fréquence aléatoire du succès. Un intervalle de fluctuation de F au seuil de 95% est un intervalle :

- de la forme $\left[\dfrac{a}{n}\,;\,\dfrac{b}{n}\right]$ où a et b sont des entiers compris entre 0 et n ;
- tel que $P\left(\dfrac{a}{n}\leq F\leq\dfrac{b}{n}\right)\geq 0,95$ ce qui équivaut à $P(a\leq X\leq b)\geq 0,95$

I. On dispose du biberon de la partie A dont la proportion de boules rouges est de 30 %.

1.a) On retourne 50 fois le biberon. Quelle est la taille de l'échantillon étudié ? On appelle X la variable aléatoire correspondant au nombre de boules rouges obtenu lors de cette expérience. Déterminer la loi de probabilité suivie par X et donner ses paramètres.

b) A l'aide du tableau situé en page 4, déterminer deux réels a et b tels que $P(a\leq X\leq b)\geq 0,95$.

2. En pratique, on détermine les entiers a et b selon le critère suivant :

a et b sont les plus petits entiers tels que $P(X\leq a)>0,025$ et $P(X\leq b)\geq 0,975$.

a) Quelles sont les valeurs de a et b qui respectent cette condition.

b) Illustrer cette situation sur les graphiques 3 et 4.

3. On appelle F la variable aléatoire qui donne la fréquence d'apparition de la boule rouge sur l'échantillon de taille 50.

a) Quelle relation avons-nous entre F et X ?

b) Donner l'intervalle de fluctuation de F correspondant à ce critère et le comparer à celui défini dans la **partie A.**

4. On renouvelle l'expérience en retournant cette fois-ci 100 fois le biberon. Reprendre les questions **2.a)** et **3.b)** avec cette nouvelle taille d'échantillon.

II. 1. On dispose à présent d'un biberon dont la proportion de boules rouges est de 10%. On renouvelle 50 fois l'expérience précédente. Quelle méthode est la plus adaptée pour obtenir l'intervalle de fluctuation ? Argumenter.

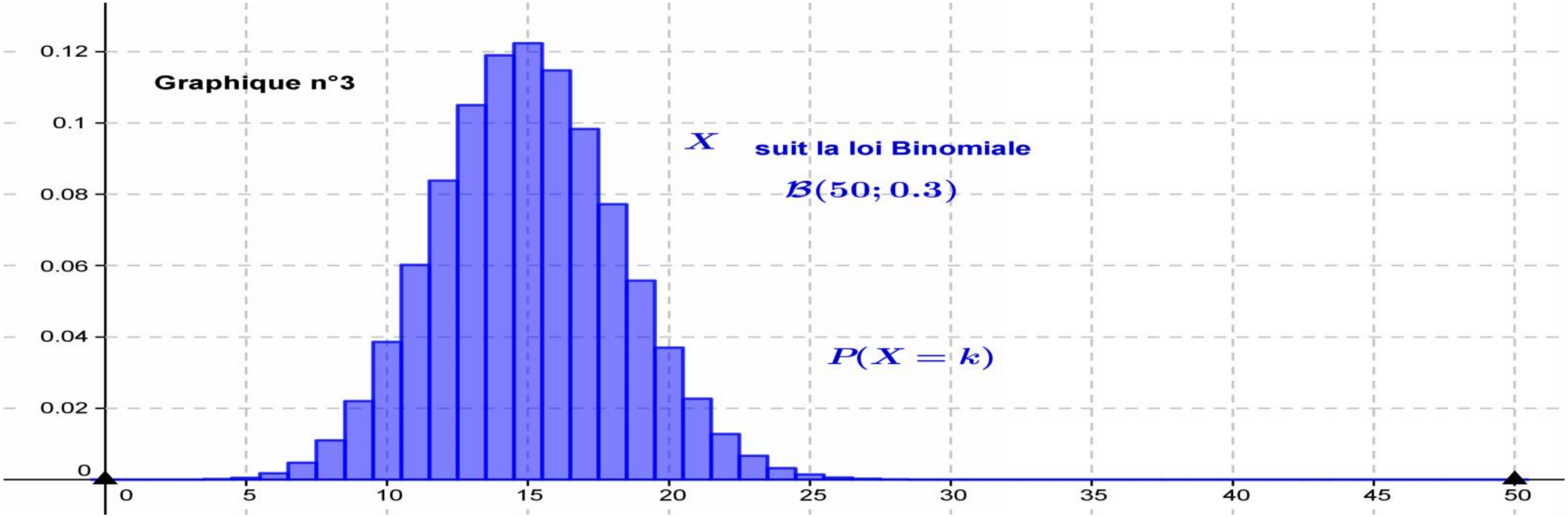
2. A l'aide de la calculatrice déterminer l'intervalle de fluctuation au seuil de 95 % de la fréquence d'apparition de la « boule rouge ».

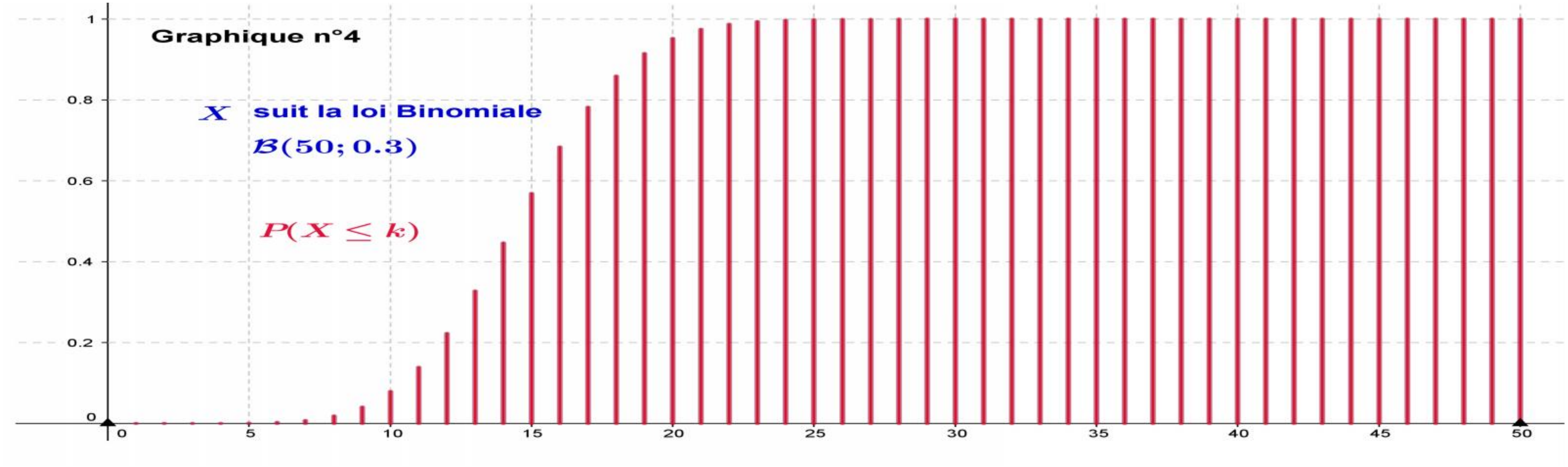
$n = 50$; $p = 0,3$

k	$P(X = k)$	$P(X \leq k)$
0	1,79847E-08	1,79847E-08
1	3,85385E-07	4,0337E-07
2	4,04655E-06	4,44992E-06
3	2,77477E-05	3,21977E-05
4	0,00013973	0,000171927
5	0,000550934	0,000722862
6	0,00177086	0,002493722
7	0,004770481	0,007264203
8	0,010989144	0,018253347
9	0,021978287	0,040231634
10	0,038618991	0,078850625
11	0,06018544	0,139036065
12	0,08382972	0,222865785
13	0,105017451	0,327883236
14	0,118948338	0,446831574
15	0,122346862	0,569178436
16	0,114700183	0,683878619
17	0,098314443	0,782193062
18	0,077247062	0,859440124
19	0,055757278	0,915197401
20	0,037038763	0,952236165
21	0,022676794	0,974912958
22	0,012810916	0,987723874
23	0,006683956	0,99440783
24	0,003222622	0,997630452
25	0,001436369	0,999066821
26	0,00059191	0,999658731
27	0,00022549	0,99988422
28	7,93815E-05	0,999963602
29	2,58088E-05	0,999989411
30	7,74263E-06	0,999997153
31	2,14082E-06	0,999999294
32	5,44762E-07	0,999999839
33	1,27347E-07	0,999999966
34	2,72886E-08	0,999999994
35	5,34635E-09	0,999999999
36	9,54705E-10	1
37	1,54817E-10	1
38	2,26987E-11	1
39	2,99324E-12	1
40	3,52775E-13	1
41	3,68754E-14	1
42	3,38652E-15	1
43	2,70021E-16	1
44	1,84105E-17	1
45	1,05203E-18	1
46	4,90076E-20	1
47	1,78751E-21	1
48	4,78798E-23	1
49	8,37548E-25	1
50	7,17898E-27	1

$n = 100$; $p = 0,3$

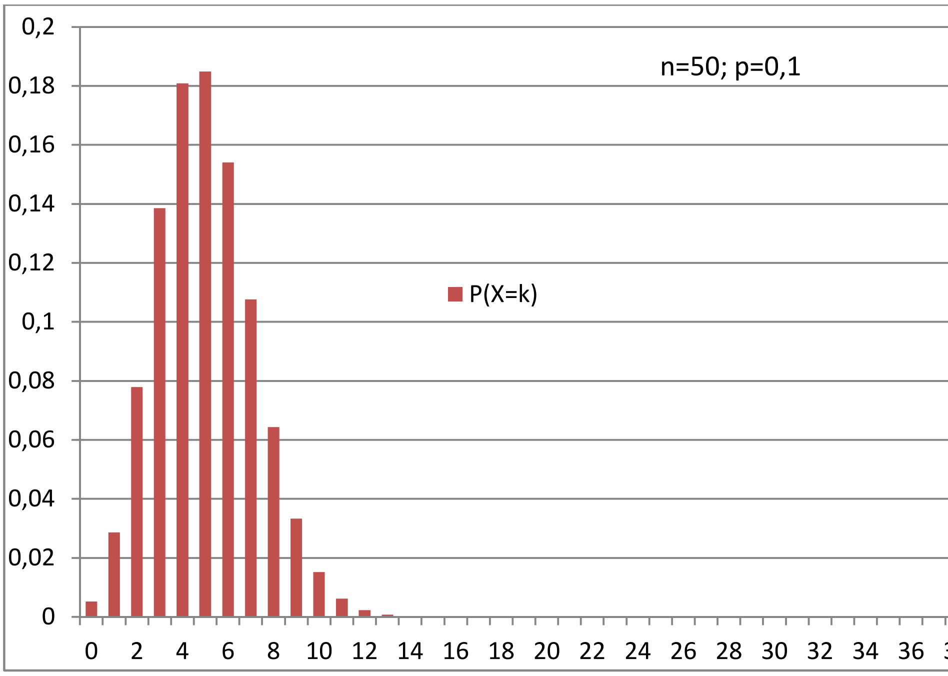
k	$P(X = k)$	$P(X \leq k)$
0	3,23448E-16	3,23448E-16
1	1,3862E-14	1,41855E-14
2	2,94073E-13	3,08259E-13
3	4,11703E-12	4,42529E-12
4	4,27877E-11	4,7213E-11
5	3,52081E-10	3,99294E-10
6	2,38912E-09	2,78842E-09
7	1,37497E-08	1,65381E-08
8	6,85027E-08	8,50408E-08
9	3,00107E-07	3,85148E-07
10	1,17042E-06	1,55557E-06
11	4,10406E-06	5,65963E-06
12	1,30451E-05	1,87047E-05
13	3,7845E-05	5,65497E-05
14	0,000100791	0,000157341
15	0,000247659	0,000405
16	0,000563866	0,000968865
17	0,001194068	0,002162933
18	0,002359706	0,004522639
19	0,004364569	0,008887208
20	0,007575645	0,016462853
21	0,0123684	0,028831253
22	0,019034486	0,047865739
23	0,027665029	0,075530767
24	0,038039414	0,113570182
25	0,049559923	0,163130104
26	0,061269135	0,22439924
27	0,071966921	0,296366161
28	0,080412019	0,376778179
29	0,085561557	0,462339736
30	0,086783865	0,549123601
31	0,083984385	0,633107986
32	0,07761057	0,710718556
33	0,068539205	0,779257761
34	0,05788395	0,837141712
35	0,046779682	0,883921394
36	0,036198564	0,920119958
37	0,026834457	0,946954414
38	0,019066588	0,966021002
39	0,012990422	0,979011424
40	0,008490169	0,987501593
41	0,005324845	0,992826437
42	0,003205774	0,996032211
43	0,001853172	0,997885383
44	0,001028871	0,998914254
45	0,000548731	0,999462985
46	0,000281182	0,999744167
47	0,000138454	0,999882622
48	6,55186E-05	0,999948141
49	2,97986E-05	0,999977939
50	1,30262E-05	0,999990965
51	5,4732E-06	0,999996439
52	2,21033E-06	0,999998649
53	8,57919E-07	0,999999507
54	3,20017E-07	0,999999827
55	1,14708E-07	0,999999941

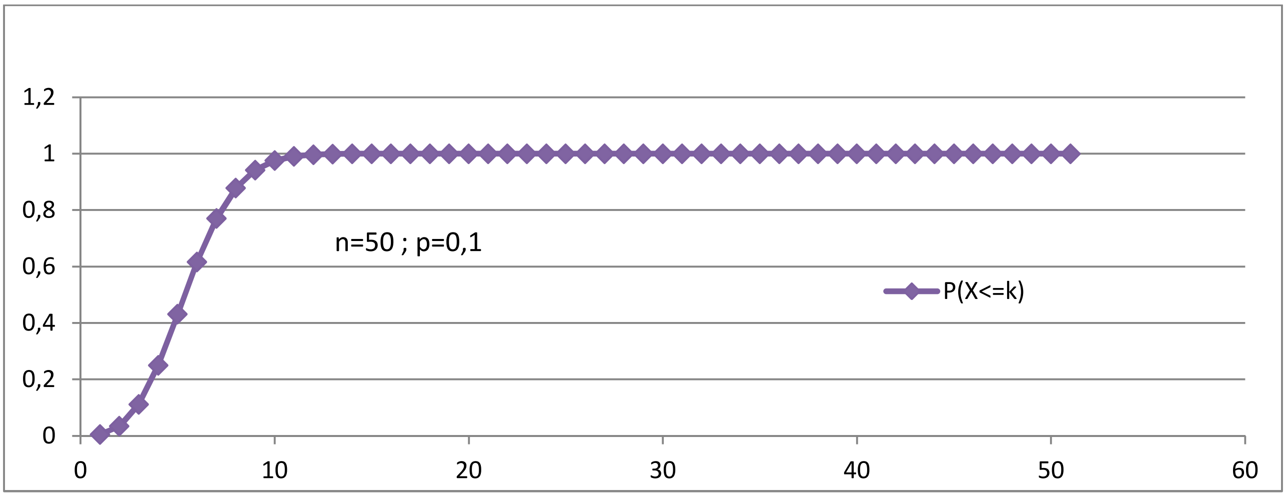

Graphique n°3
X suit la loi Binomiale
$\mathcal{B}(50; 0.3)$
$P(X = k)$
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
35
40
45
50


Graphique n°4
X suit la loi Binomiale
$\mathcal{B}(50; 0.3)$
$P(X \leq k)$
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
35
40
45
50

Annexe 3 : $n = 50$; $p = 0{,}1$

k	$P(X = k)$	$P(X \le k)$
0	0,005153775	0,005153775
1	0,028632084	0,03378586
2	0,077942897	0,111728756
3	0,13856515	0,250293906
4	0,180904501	0,431198407
5	0,184924601	0,616123008
6	0,154103834	0,770226842
7	0,107628075	0,877854916
8	0,064277878	0,942132794
9	0,03332927	0,975462064
10	0,015183334	0,990645398
11	0,00613468	0,996780079
12	0,002215301	0,99899538
13	0,0007195	0,99971488
14	0,000211282	0,999926161
15	5,63418E-05	0,999982503
16	1,36942E-05	0,999996197
17	3,04315E-06	0,99999924
18	6,19901E-07	0,99999986
19	1,16005E-07	0,999999976
20	1,99786E-08	0,999999996
21	3,17121E-09	0,999999999
22	4,6447E-10	1





C. Utilisation de l'intervalle de fluctuation pour la prise de décision.

I. On a retrouvé un biberon utilisé en classe de seconde, en revanche on a un doute sur la proportion de boules rouges contenues dans le biberon. On suppose alors que cette proportion est de 0,3.

1. Rappeler l'intervalle de fluctuation au seuil de 0,95 en utilisant la loi Binomiale avec $n = 100$.

2. Un élève a retourné 100 fois ce biberon et a obtenu 0,33 comme fréquence d'apparition de boules rouges. Est-ce raisonnable de penser qu'il y a 3 boules rouges parmi les 10 dans ce biberon ?

II. 1. En fait, l'élève croit se souvenir qu'il y avait placé 4 boules rouges parmi les 10. Que penser de cette hypothèse ? Argumenter.

2. On renouvelle l'expérience précédente sur un échantillon de taille 500 et on obtient 0,328 comme fréquence d'apparition de boules rouges. Peut-on à présent conclure ?