Corrigé de lœxercice n° 2 du DM

Question n°1

La loi qui compte le nombre dœissues favorables de lœévénement S suite à 50 lancers indépendants est la loi binomiale de paramètre : n=50 et p=P(S)=0,375

On note cette loi B (50, 0.375)

Remarque : on peut vérifier le nombre 0,375 en calculant le nombre de combinaisons favorables par rapport au nombre de combinaisons possibles et on obtient bien la probabilité $\frac{4\times3\times2}{4\times4\times4}=\frac{3}{8}=0,375$

Question n°2

Si la variable aléatoire X suit la loi binomiale B (n, p), alors dœaprès le cours on sait que :

$E(X) = n\times p$ et $V(X) = n\times p\times(1-p)$ avec dans cet exercice $n=50$ et $p=0,375$

Réponses : $E(X) \approx 18,76$ et $V(X) \approx 11,72$ et $\sigma=\sqrt{V(x)}=\sqrt{n\times p\times(1-p)}\approx 3,42$

Question n°3.a

Le graphique représente le calcul de $P(X\le k)$ avec k un entier naturel tel que $0\le k\le 50$

La fonction F définie par $F(k)=P(X\le k)$ avec k un entier naturel sœappelle la fonction de répartition de la variable X de loi binomiale $\mathcal{B}$ (50 , 0.375)

Réponse: Lecture sur le graphique $P(X\le 20) \approx 0,7$ *(70% de chance que S se réalise au plus 20 fois)*

Question n°3.b

Soit a et b 2 entiers naturels donnés tels que $a<b$

$$P(a\le X\le b)=P\big((a\le X)\cap(X\le b)\big)=P(a\le X)+P(X\le b)-P\big((a\le X)\cup(X\le b)\big)$$

1) Calcul de $P\big((a\le X)\cup(X\le b)\big)$

Lœévénement $(a\le X)\cup(X\le b) \Leftrightarrow (a\le X)$ ou $(X\le b) \Leftrightarrow 0\le X\le 50$ car $a<b$

donc $P\big((a\le X)\cup(X\le b)\big)=1$

2) Calcul de $P(a\le X)$: D'après la formule sur **l'événement contraire** on a : $P(a\le X)=1-P(X<a)$

Comme $X<a \Leftrightarrow X\le P_E(a)-1$ a avec $P_E(a)=\text{partie entiere de a}$

EXPLICATIONS : Comme la variable X **ne peut prendre que des valeurs entières** entre 0 et 50

on a $X<a \Leftrightarrow X\le a-1$ a (par exemple $X<15 \Leftrightarrow X\le 15-1 \Leftrightarrow X\le 14$)

En généralisant on obtient que $P(a\le X) = 1-P(X\le a-1)$

Réponse : $\boxed{P(15\le X\le 20) = P(X\le 20) - P(X\le 14)}$

$P(a\le X\le b)=P(a\le X)+P(X\le b)-P\big((a\le X)\cup(X\le b)\big)=\left[1-P(X\le a-1)\right]+P(X\le b)-1=P(X\le b)-P(X\le a-1$

Donc $\boxed{P(15\le X\le 20) = P(X\le 20) - P(X\le 14)\approx 0,7-0,1 \approx 0,6}$

Question n°3.c

On cherche lœévénement E tel $P(E)\ge 0,5$

Dœaprès la représentation graphique on constate que :

$P(X\le 18)< 0,5$ et $P(X\le 19)>0,5$

Réponse: Le jeu S se réalise dans plus de 50% des cas **si** $\mathbf{X\ge 19}$

cœest-à-dire : il faut que S se réalise au moins 19 fois (19 succès)

Question n°3.d

On cherche **la plus petite valeur a** telle que $P(X \le a) > 0{,}025$

Réponse : D'après le fichier Excel qui simule la fonction de répartition du nombre de succès du jeu S sur 50 lancers (voir annexe ci-dessous), on trouve que $\mathbf{a = 12}$

Question n°3.e

On cherche **la plus petite valeur b** telle que $P(X \le b) \ge 0{,}975$

Réponse : D'après le fichier Excel qui simule la fonction de répartition du nombre de succès du jeu S sur 50 lancers (voir annexe ci-dessous), on trouve que $\mathbf{b = 26}$

Question n°4

D'après les 2 questions précédentes on a : $\left[\dfrac{a}{n} ; \dfrac{b}{n}\right] = \left[\dfrac{12}{50} ; \dfrac{26}{50}\right] = [\,0.24 \;;\; 0.52\,]$

Réponse : On rejette l'hypothèse $P(S) = 0{,}375$ pour la machine $n°3$ ***(avec marge d'erreur de 5%)***

Annexe : Fichier Excel (LES 2 représentations graphiques d'une LOI BINOMIALE)

CI-DESSOUS : Représentation graphique de la fonction F définie par : $F(k) = P(X \le k)$

et qui est appelée **la fonction de répartition de la variable aléatoire X**

La variable aléatoire X suit la loi BINOMIALE de paramètre n = 50 et p = 0,375

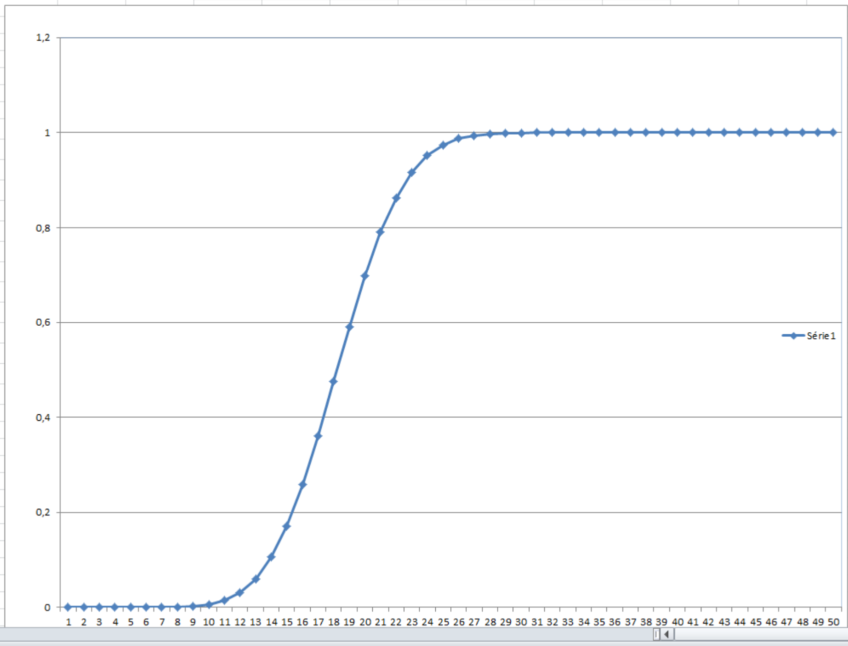
donc pour tout entier naturel k tel que $0 \le k \le 50$ on a la formule :

$$F(k) = P(X \le k) = P(X=0) + P(X=1) + P(X=2) + \ldots\ldots + P(X \le k-1) + P(X=k)$$

B12 =LOI.BINOMIALE(A12;50;0,375;1)

	A	B	C	D
1	1	1,93E-09		Tableau sur la loi Binomiale de paramètre n=50 et p=0,375
2	2	2,94E-08		
3	3	2,93E-07		
4	4	2,15E-06		
5	5	1,24E-05		
6	6	5,85E-05		
7	7	0,000233		
8	8	0,000794		
9	9	0,002365		
10	10	0,00623		
11	11	0,014664		
12	**12**	**0,031109**	**a**	*P (X <= a) > 0.025*
13	13	0,05995		
14	14	0,105686		
15	15	0,171544		
16	16	0,257984		
17	17	0,361711		
18	18	0,475811		
19	19	0,591112		
20	20	0,698342		
21	21	0,790253		
22	22	0,862947		
23	23	0,916044		
24	24	0,951886		
25	**25**	**0,97425**	**b**	*P (X <= b) >= 0.9725*
26	26	0,987153		
27	27	0,994035		
28	28	0,997426		
29	29	0,99897		
30	30	0,999618		
31	31	0,999869		
32	32	0,999959		
33	33	0,999988		
34	34	0,999997		
35	35	0,999999		
36	36	1		
37	37	1		
38	38	1		

Ce grahique représente P(X <=k) avec k = 0 , 1 , 2 ,3 , ... , 48 , 49 , 50 *(colonne B)*



Feuil2 Feuil3

Prêt

Commentaires sur le résultat obtenu :

En classe de seconde, le cours de mathématiques introduit l'intervalle :

$$\left[p - \frac{1}{\sqrt{n}} \;;\; p + \frac{1}{\sqrt{n}}\right] = \left[p - \frac{1}{\sqrt{50}} \;;\; p + \frac{1}{\sqrt{50}}\right] \approx [\,0{,}234 \;;\; 0{,}516] \quad \text{car } p = P(S) = 0{,}375$$

Rappel du cours de la classe de seconde : INTERVALLE DE FLUCTUATION au seuil de 95%

Théorème 1 :

Soit un échantillon de taille $n \geqslant 25$ sur lequel on observe un caractère de probabilité d'apparition p tel que $0,2 \leqslant p \leqslant 0,8$.
Si f désigne la fréquence du caractère observé dans l'échantillon, alors f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}} \; ; \; p + \frac{1}{\sqrt{n}} \right]$ avec une probabilité d'au moins $0,95$.
Cet intervalle est appelé ***intervalle de fluctuation de taille n au seuil de 95%****.*

Application 1 : Prévision

Autrement dit, dans au moins 95% des cas, l'écart entre la fréquence observée et la probabilité est inférieur à $\frac{1}{\sqrt{n}}$. Ainsi, si je réalise un échantillon, je peux affirmer en prenant un risque de 5%, que la fréquence que j'obtiendrais sera dans l'intervalle de fluctuation.
La valeur $\frac{1}{\sqrt{n}}$ est appelé la ***marge d'erreur au seuil de 95%****.*

Application 2 : Prise de décision

Dans une population donnée, on étudie un caractère et on émet l'hypothèse :

La proportion du caractère est p

Pour juger de cette hyptothèse, on prélève un échantillon de taille n, et on calcule la fréquence, notée f, du caractère au sein de l'échantillon.

- *Si $f = p$, on ne rejette pas l'hyptothèse mais ce cas est peu fréquent en raison de la fluctuation d'échantillonnage.*
- *Si $f \neq p$, on regarde l'écart entre f et p :*
  - ◇ *Si cet écart (la marge d'erreur) est inférieur à $\frac{1}{\sqrt{n}}$ (i.e f est dans l'intervalle de fluctuation), on ne rejette pas l'hypothèse au motif que le hasard produirait un tel écart dans 95% des échantillons envisageables.*
  - ◇ *Si cet écart est supérieur à $\frac{1}{\sqrt{n}}$ (i.e f est hors l'intervalle de fluctuation), on rejette l'hyptothèse puisqu'un tel écart n'apparait que dans 5% des échantillons envisageables. Soit il s'agit vraiment un échantillon exceptionnel (et on a eu tort de rejeter l'hyptothèse) soit l'intervalle de fluctuation n'est pas correct, ce qui signifie que la proportion choisie n'était pas adaptée.*

Application 3 : Estimation

On adopte la procédure d'estimation suivante :
Pour estimer la probabilité d'un caractère au sein d'une population, on prélève UN échantillon aléatoire de taille n pour lequel on obtient UNE fréquence f. La probabilité p, inconnue, de la population mère est dans l'intervalle $\left[f - \frac{1}{\sqrt{n}} \; ; \; f + \frac{1}{\sqrt{n}} \right]$ dans 95% des cas. Cet intervalle est appelé ***intervalle de confiance de p au niveau de confiance 95%****.*

<u>A retenir par cœur</u> :

1) L'intervalle $\left[p-\frac{1}{\sqrt{n}} \ ; \ p+\frac{1}{\sqrt{n}} \right]$ est appelé : intervalle de fluctuation de taille n au seuil de 95%

2) L'intervalle $\left[f-\frac{1}{\sqrt{n}} \ ; \ f+\frac{1}{\sqrt{n}} \right]$ est appelé : intervalle de confiance avec taux de de confiance de 95%

avec p la proportion d'un caractère (ou la probabilité) au niveau de la population « mère »

et f la fréquence « observée » (ou la probabilité) de ce caractère dans un échantillon donné de taille n

Calculons dans cet exercice l'intervalle de fluctuation au seuil de 95% avec cette formule :

$$\left[p-\frac{1}{\sqrt{n}} \ ; \ p+\frac{1}{\sqrt{n}} \right]=\left[0,375-\frac{1}{\sqrt{50}} \ ; \ 0,375+\frac{1}{\sqrt{50}} \right]\approx\left[0,234 \ ; \ 0,516\right]$$

On constate qu'on a: $\left[\frac{a}{n};\frac{b}{n} \right]\subset\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}} \right]$

En effet on a : $\left[\frac{a}{n};\frac{b}{n} \right]=\left[0.24 \ ; \ 0.52 \right]\subset\left[0,375-\frac{1}{\sqrt{50}} \ ; \ 0,375+\frac{1}{\sqrt{50}} \right]\approx\left[0,234 \ ; \ 0,516\right]$

Ce qui justifie la formule simplifiée sur l'intervalle de fluctuation qu'on apprend en classe de seconde .

<u>Ce qu'il faut retenir de ce DM</u>

Ce qu'il faut retenir : $\left[\frac{a}{n};\frac{b}{n} \right]=\left[\frac{12}{50};\frac{26}{50} \right]=\left[0.24 \ ; \ 0.53 \right]$ s'appelle intervalle de fluctuation de taille n au seuil de 95% (

ET $P\left(\frac{a}{n}\le\frac{X}{n}\le\frac{b}{n}\right)=P\left(\frac{X}{n}\le\frac{b}{n}\right)-P\left(\frac{X}{n}<\frac{a}{n}\right)\ge 0.975-0.025=0.95$ (c'est-à-dire 95%)

<u>Conclusion</u> : Prise de décision suite à ***n*** tests sur une machine donnée (hypothèse : « $P(S)=0.375$ »)

Si ***f*** est la fréquence du nombre de succès du jeu S après ***n*** lancers **sur une machine donnée** alors :

1. Si $f\in\left[\frac{a}{n};\frac{b}{n} \right]$ alors on aura tendance à croire que la machine est **équilibrée**

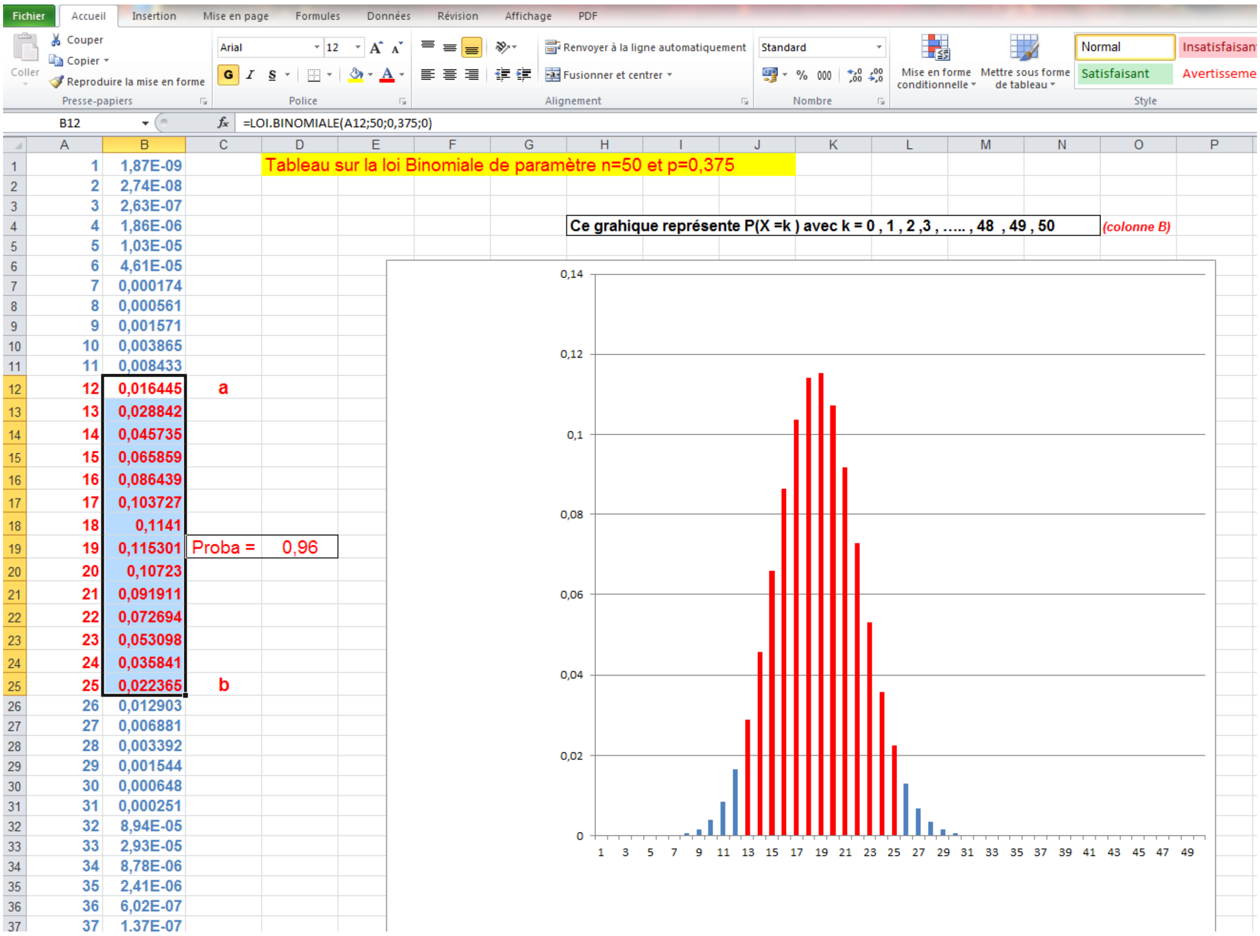
et on ne rejette pas l'hypothèse « $P(S)=0.375$ » *avec une marge d'erreur de 5%....*

2. Si $f\notin\left[\frac{a}{n};\frac{b}{n} \right]$ alors on aura tendance à croire que la machine est **NON équilibrée**

et on rejette l'hypothèse « $P(S)=0.375$ » *à tort dans* ***5%*** *des cas* .

Annexe : Calcul de la probabilité sur l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]=\left[\frac{12}{50};\frac{25}{50}\right]$

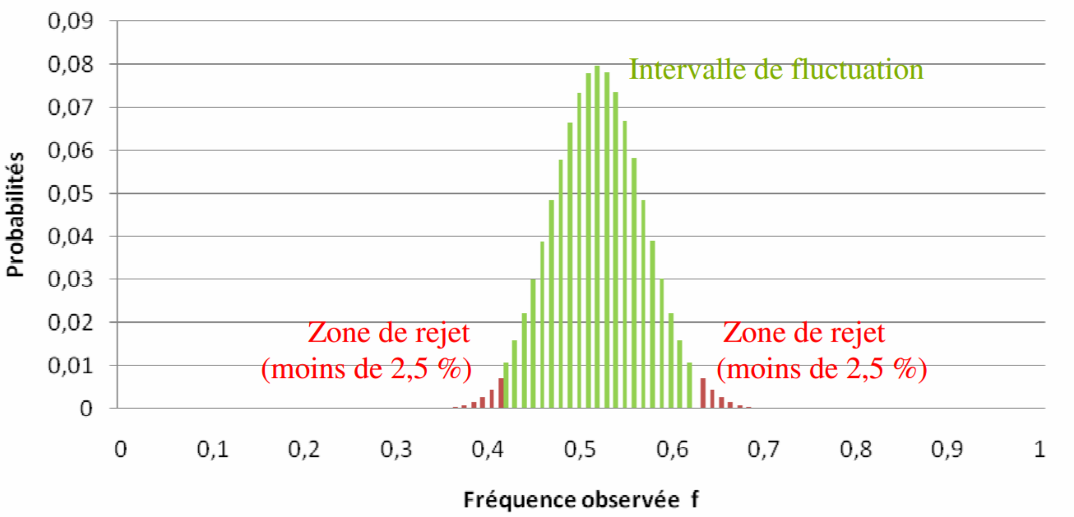
$$P\left(\frac{a}{n}\le X\le\frac{b}{n}\right)=0.96\ge 0.975-0.025=0.95$$



	A	B	C	D
1	1	1,87E-09		
2	2	2,74E-08		
3	3	2,63E-07		
4	4	1,86E-06		
5	5	1,03E-05		
6	6	4,61E-05		
7	7	0,000174		
8	8	0,000561		
9	9	0,001571		
10	10	0,003865		
11	11	0,008433		
12	12	0,016445	a	
13	13	0,028842		
14	14	0,045735		
15	15	0,065859		
16	16	0,086439		
17	17	0,103727		
18	18	0,1141		
19	19	0,115301	Proba =	0,96
20	20	0,10723		
21	21	0,091911		
22	22	0,072694		
23	23	0,053098		
24	24	0,035841		
25	25	0,022365	b	
26	26	0,012903		
27	27	0,006881		
28	28	0,003392		
29	29	0,001544		
30	30	0,000648		
31	31	0,000251		
32	32	8,94E-05		
33	33	2,93E-05		
34	34	8,78E-06		
35	35	2,41E-06		
36	36	6,02E-07		
37	37	1,37E-07		

Autre exemple de calcul d'un intervalle de fluctuation de taille n=100 au seuil de 95%

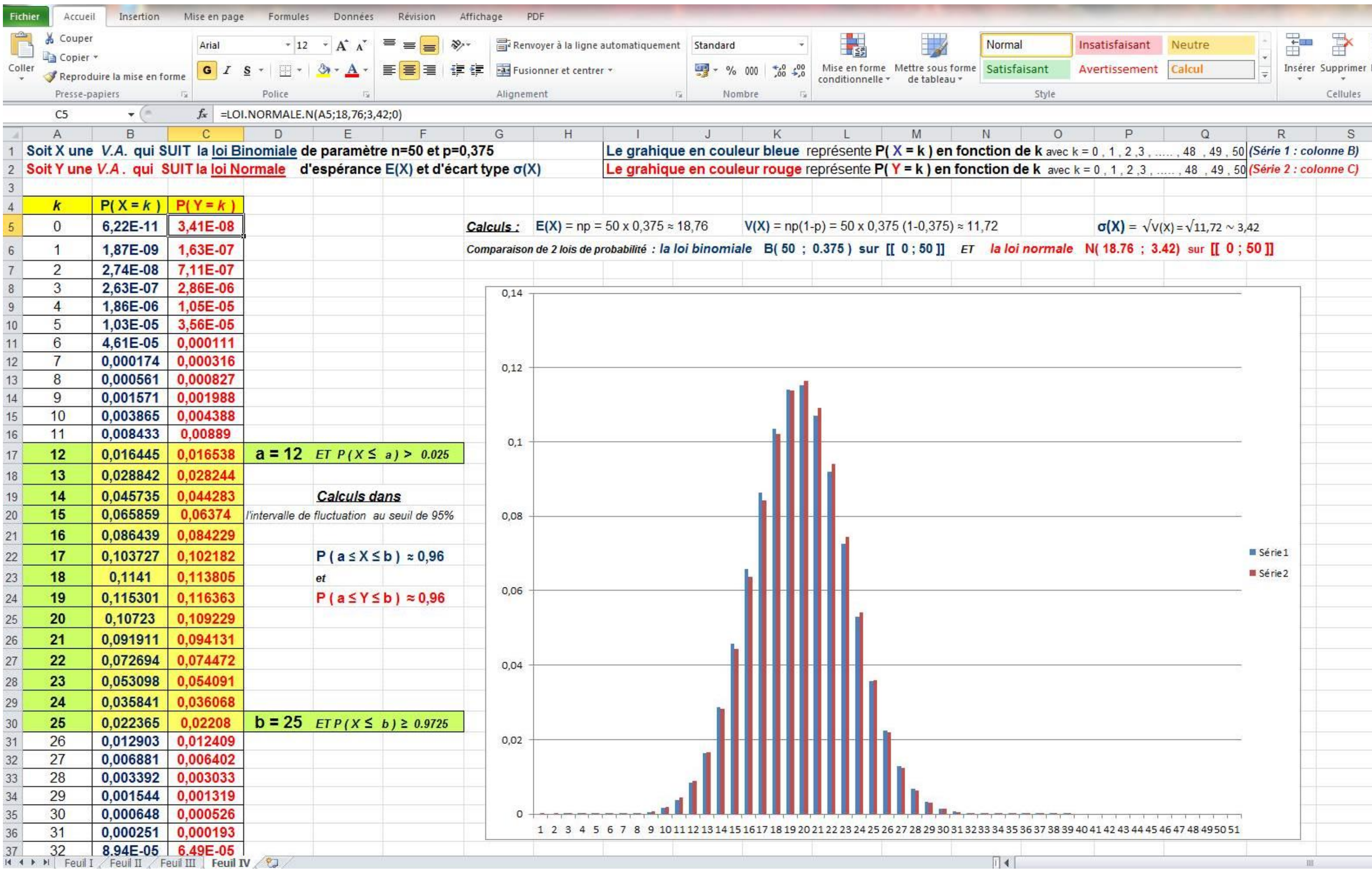
Remarque : la recherche de l'intervalle de fluctuation peut-être illustrée par le diagramme en bâton de la loi binomiale de paramètres $n = 100$ et $p = 0,52$.



ANNEXE : Le théorème de **Moivre Laplace** **(Niveau classe Terminale)**

Rappel : En probabilités, selon le théorème de De Moivre-Laplace :
si la variable X_n suit une loi binomiale de paramètre n et p tel que $p \in \left]0\,;1\right[$ alors la variable

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$$ converge en loi vers une loi normale centrée et réduite, qu'on note $N(0,1)$:



k	P(X = k)	P(Y = k)
0	6,22E-11	3,41E-08
1	1,87E-09	1,63E-07
2	2,74E-08	7,11E-07
3	2,63E-07	2,86E-06
4	1,86E-06	1,05E-05
5	1,03E-05	3,56E-05
6	4,61E-05	0,000111
7	0,000174	0,000316
8	0,000561	0,000827
9	0,001571	0,001988
10	0,003865	0,004388
11	0,008433	0,00889
12	0,016445	0,016538
13	0,028842	0,028244
14	0,045735	0,044283
15	0,065859	0,06374
16	0,086439	0,084229
17	0,103727	0,102182
18	0,1141	0,113805
19	0,115301	0,116363
20	0,10723	0,109229
21	0,091911	0,094131
22	0,072694	0,074472
23	0,053098	0,054091
24	0,035841	0,036068
25	0,022365	0,02208
26	0,012903	0,012409
27	0,006881	0,006402
28	0,003392	0,003033
29	0,001544	0,001319
30	0,000648	0,000526
31	0,000251	0,000193
32	8.94E-05	6.49E-05

La variable aléatoire $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ **est une variable aléatoire CENTREE et REDUITE car**

la variable aléatoire X_n suit la loi binomiale B (n, p), alors d'après le cours on sait que :

$$E(X_n) = n \times p \quad \text{et} \quad V(X_n) = n \times p \times (1-p) \quad \text{et} \quad \sigma_{X_n} = \sqrt{V(X_n)}$$

donc on a : $Z_n = \dfrac{X_n - E(X_n)}{\sigma_{X_n}}$