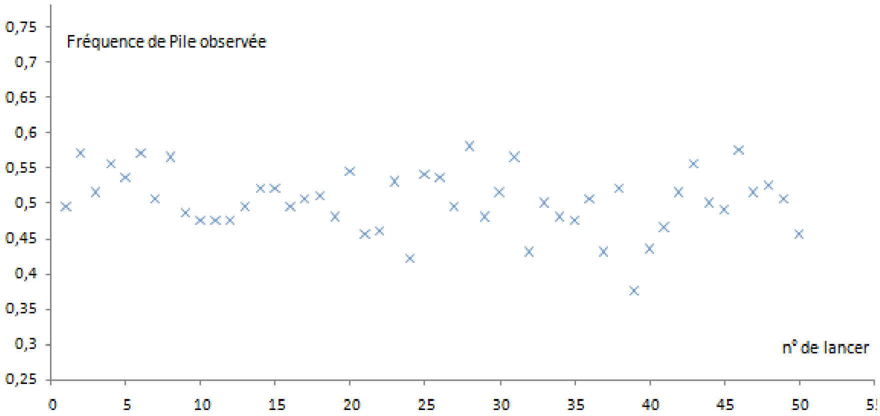
Echantillonnage.

I. Intervalle de fluctuation et intervalle de confiance.

On considère l'expérience aléatoire suivante : « On lance une pièce 200 fois et on calcule la fréquence d'obtention de Pile ». On demande à 50 personnes de réaliser cette expérience.

Par exemple, le premier lanceur obtient 99 fois Pile. La fréquence d'obtention de Pile qu'il a observé est donc f =...........
On a représenté les résultats obtenus par les 50 lanceurs sur le dessin ci contre (simulation faite avec un tableur)
On remarque que ....
....
....
....



Propriété (admise)

On considère une population dont une proportion p des individus possède un caractère donné. On prélève dans cette population un échantillon de taille n. Si 0,2≤ p ≤0,8 et avec n ≥ 25 alors dans 95% des cas, la fréquence f du caractère dans l'échantillon appartient à l'intervalle. $p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}$.
Cet intervalle s'appelle l'intervalle de fluctuation au seuil de 95% de la fréquence f.

Dans l'expérience ci-dessus, la proportion d'obtention de Pile .... *donc pour un échantillon de taille* ......
l'intervalle de fluctuation au seuil de 95% de la fréquence d'obtention de Pile observés est ....
Soit.................. . *Ce qui signifie que dans 95% des cas, la fréquence de Pile obtenus appartiendra à* ....
Ce qui signifie que dans 95% des cas, le nombre de Pile obtenus appartiendra à ....
....

☑ Savoir faire : Savoir déterminer un intervalle de fluctuation au seuil de 95 %:

Lors d'une élection, 60% des électeurs ont voté pour le candidat A. On interroge au hasard à la sortie des urnes 200 personnes. On compte le nombre de personnes qui ont voté pour le candidat A. Déterminer l'intervalle de fluctuation au seuil de 95% de la fréquence des personnes ayant voté pour le candidat A.
....
....
....

Propriété (admise)

On considère une population dont une proportion p des individus possède un caractère donné. Soit f la fréquence observée pour un échantillon de taille n. Si 0,2≤ p ≤0,8 et avec n ≥25 alors dans 95% des cas, la proportion p appartient à l'intervalle. $f-\frac{1}{\sqrt{n}}\,;f+\frac{1}{\sqrt{n}}$. Cet intervalle s'appelle l'intervalle de confiance de p.

☑ Savoir faire : Savoir utiliser un intervalle de confiance :

Lors d'une élection, on interroge au hasard à la sortie des urnes 200 personnes. On compte le nombre de personnes qui ont voté pour le candidat A. On en compte 121. Un sondage national prévoyait 70% d'intention de vote pour le candidat A. Est-ce cohérent ?
....
....
....
....

II. Intervalle de fluctuation et loi binomiale.

Propriété

On considère une population dont une proportion p des individus possède un caractère donné. On prélève dans cette population un échantillon de taille n. La variable aléatoire qui compte le nombre d'individus possédant ce caractère suit la loi binomiale $B(n , p)$.

Exemple :

On suppose que 45% des français sont propriétaires de leur logement. On interroge au hasard 50 personnes. Soit X la variable aléatoire qui compte le nombre k de personnes propriétaires de leur logement. La variable aléatoire X suit la loi binomiale $B(50 , 0{,}45)$.Avec un tableur on obtient

H6 =LOI.BINOMIALE(H5;50;0,45;0)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1																											
2	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
3	P(X=k)	1E-13	4E-12	9E-11	1E-09	1E-08	8E-08	5E-07	3E-06	1E-05	4E-05	1E-04	4E-04	0,001	0,003	0,006	0,012	0,021	0,034	0,051	0,07	0,089	0,104	0,112	0,112	0,103	0,087
4																											
5	k	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	
6	P(X=k)	0,069	0,05	0,034	0,021	0,012	0,006	0,003	0,001	6E-04	2E-04	7E-05	2E-05	6E-06	2E-06	3E-07	7E-08	1E-08	2E-09	2E-10	3E-11	2E-12	2E-13	8E-15	3E-16	5E-18	
7																											

Pour k <...... et k >......, les probabilités sont inférieures à 10^{-3} et peuvent être considérées comme négligeables.

Définition

On considère une population dont une proportion p des individus possède un caractère donné. On prélève dans cette population un échantillon de taille n. Soit X la variable aléatoire associée au nombre d'individus possédant ce caractère. Soit a et b les plus petits entiers tel que $P(X\leq a) > 0{,}025$ et $P(X\leq b) \geq 0{,}975$.

L'<u>intervalle de fluctuation au seuil de 95 %</u> associée à la variable aléatoire X est : $\frac{a}{n} ; \frac{b}{n}$.

On peut utiliser un tableur pour calculer les probabilités cumulées, on obtient :

G8 =LOI.BINOMIALE(G6;50;0,45;1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1																												
2	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
3	P(X=k)	1E-13	4E-12	9E-11	1E-09	1E-08	8E-08	5E-07	3E-06	1E-05	4E-05	1E-04	4E-04	0,001	0,003	0,006	0,012	0,021	0,034	0,051	0,07	0,089	0,104	0,112	0,112	0,103	0,087	
4	P(X≤k)	1E-13	4E-12	9E-11	1E-09	1E-08	9E-08	6E-07	3E-06	1E-05	6E-05	2E-04	6E-04	0,002	0,004	0,01	0,022	0,043	0,077	0,127	0,197	0,286	0,39	0,502	0,613	0,716	0,803	
5																												
6	k	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50		
7	P(X=k)	0,069	0,05	0,034	0,021	0,012	0,006	0,003	0,001	6E-04	2E-04	7E-05	2E-05	6E-06	2E-06	3E-07	7E-08	1E-08	2E-09	2E-10	3E-11	2E-12	2E-13	8E-15	3E-16	5E-18		
8	P(X≤k)	0,872	0,922	0,956	0,976	0,988	0,995	0,998	0,999	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
9																												

On lit alors que a = ... et b = ...On en déduit que l'intervalle de fluctuation au seuil de 95% est

.... Soit.................. . Ce qui signifie que ....

....

Remarque : Avec la formule de seconde on obtient comme intervalle de fluctuation au seuil de 95 %

....

....

☑ Savoir faire : Savoir déterminer un intervalle de fluctuation au seuil de 95 %:

Une urne contient 60% de boules blanches. On tire 100 boules avec remise. Déterminer l'intervalle de fluctuation au seuil de 95% de la fréquence observée d'une boule blanche.

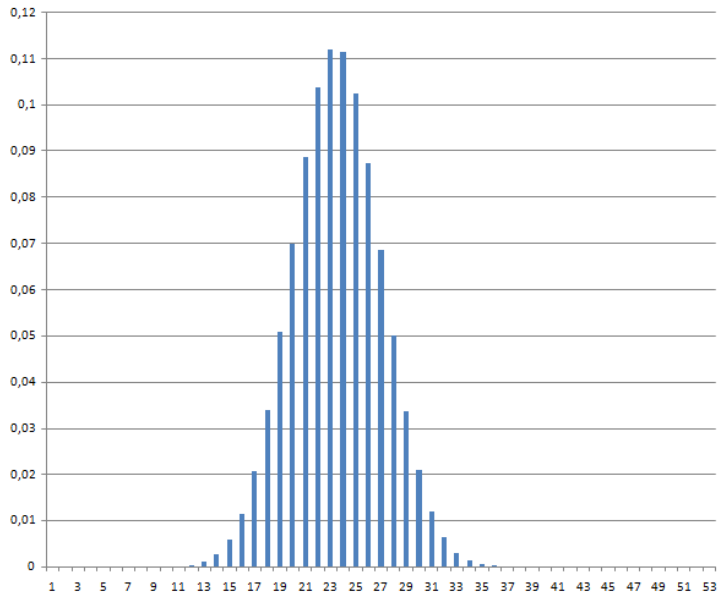
....

....

....

	Casio	Texas	Open Office	Excel
Syntaxe	Touche OPTN, puis choisir *STAT*, puis *DIST*, puis *BINM*, puis *Bpd* ou *Bcd* (voir p. 278).	Menu *distrib* (2nde var), puis choisir *binomFdp(* ou *binomFrép(*, voir p. 274.	Fonction **LOI.BINOMIALE**	
$P(X=k)$	BinominalPD(k,n,p)	binomFdp(n,p,k)	=LOI.BINOMIALE (k;n;p;0)	=LOI.BINOMIALE (k;n;p;FAUX)
$P(X\leqslant k)$	BinominalCD(k,n,p)	binomFRép(n,p,k)	=LOI.BINOMIALE (k;n;p;1)	=LOI.BINOMIALE (k;n;p;VRAI)

Remarque :
On a représenté ci contre graphiquement la loi binomiale de l'exemple précédent.
…………………………………………………………………………
…………………………………………………………………………
…………………………………………………………………………
…………………………………………………………………………
…………………………………………………………………………
…………………………………………………………………………
…………………………………………………………………………
…………………………………………………………………………



III. Prise de décision.

Définition

Si, sur un échantillon prélevé, la fréquence observée appartient à l'intervalle de fluctuation, on accepte l'hypothèse au seuil de 95 %. Dans le cas contraire, on la rejette.

Exemple :
On reprend l'exemple du paragraphe précédent. On interroge 50 personnes. Parmi ceux-là, 21 sont propriétaires de leur logement. Peut-on accepter l'hypothèse que 45% des français sont propriétaires de leur logement? L'intervalle de fluctuation au seuil de 95 % est…………………
La fréquence observée est égale à ……………………………………
……
……
……

☑ *Savoir faire : Savoir accepter ou rejeter une hypothèse :*
Un laboratoire annonce que 40% des fumeurs qui ont consommé sa pilule magique arrêtent de fumer.
Pour tester cette affirmation, on essaye sur 100 fumeurs. Soit X le nombre de personnes testées qui arrêtent de fumer.
1) Quelle loi suit X ?
……
……
2) Déterminer l'intervalle de fluctuation au seuil de 95 %.
……
……
……
……
3) Sur les 100 fumeurs testés, 30 ont arrêté. Que peut-on dire de l'annonce faite par le laboratoire ?
……
……
……
……

IV. Intervalle de fluctuation asymptotique.

1) définition.

Propriété

Soit X_n une variable aléatoire qui suit une loi binomiale $B(n\ ,\ p)$. La probabilité que la fréquence $F_n = \dfrac{X_n}{n}$ prenne ses valeurs dans l'intervalle $I_n = \left[p - 1{,}96 \times \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1{,}96 \times \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ se rapproche de 0,95 quand la taille de l'échantillon n devient grande. On appelle cet intervalle un intervalle de fluctuation asymptotique au seuil de 95% de la variable aléatoire fréquence $F_n = \dfrac{X_n}{n}$.

Remarque :

♦ *On admet que l'on peut utiliser cet intervalle lorsque n⁄30, np⁄5 et n(1- p)⁄5.*

♦ le 1,96 dans la définition de I_n *nous rappelle* ……………………………………………………………………

………

☑ *Savoir faire : Savoir déterminer un intervalle de fluctuation asymptotique au seuil de 95% :*

Une enquête a révélé que 40% des ventes d'un jouet sont d'une marque A. Un petit magasin prévoit la vente de 30 jouets. Une grande surface prévoit la vente de 1000 jouets. On désire connaitre, avec une probabilité de 0,95 le nombre de jouets de la marque A vendus dans les deux cas.

Pour chacun des deux cas, calculer l'intervalle de fluctuation asymptotique au seuil de 95% de la fréquence de jouets de la marque A vendu : I_{30} et I_{100}. Comparer l'amplitude des intervalles.

………

………

………

………

………

………

………

………

☑ *Savoir faire : Savoir accepter ou rejeter une hypothèse :*

Une marque de bonbons vend des paquets constitués de bonbons de cinq couleurs différentes, dans des proportions affichées : 20% sont marrons, 20% jaunes,10% rouges, 30% bleus et 20% verts.

Les élèves d'une classe de terminale ont voulu vérifier ces informations. Ils ont choisi d'observer un échantillon aléatoire de bonbons de cette marque. Sur 690 bonbons, ils ont dénombré 140 bonbons marrons, 152 jaunes et 125 rouges.

1) Déterminer l'intervalle de fluctuation asymptotique J au seuil de 95 % pour les bonbons marrons.

………

2) Calculer la proportion de bonbons marrons dans l'échantillon, que peut-on conclure ?

………

………

3) faire de même pour les bonbons jaunes et rouges.

………

………

………

………

………

………

V. Estimation.

Dans ce paragraphe, on suppose que la proportion *p* du caractère étudié est inconnue.
C'est le problème inverse de celui de l'échantillonnage. A partir de la fréquence observée sur un échantillon, on va estimer la proportion *p* d'un caractère dans la population tout entière.

Propriété

Soit X_n une variable aléatoire qui suit une loi binomiale $B(n\ ,\ p)$. $F_n = \frac{X_n}{n}$ *est la fréquence associée à* X_n. Pour n suffisamment grand, *p* appartient à l'intervalle $I_n = \left[Fn - \frac{1}{\sqrt{n}}\ ;\ Fn + \frac{1}{\sqrt{n}} \right]$ avec une probabilité supérieure ou égale à 0,95. Soit *f* une fréquence observée du caractère étudié sur un échantillon de taille *n*, on appelle intervalle de confiance de la proportion *p* au niveau de confiance 0,95, l'intervalle $\left[f - \frac{1}{\sqrt{n}}\ ;\ f + \frac{1}{\sqrt{n}} \right]$.

Remarque :
♦ Un niveau de confiance 0,95 signifie que dans 95 cas sur 100, on affirme à juste titre que p appartient à l'intervalle de confiance.
♦ Il n'est pas vrai d'affirmer que p est égal au centre de l'intervalle de confiance. Il n'est pas possible d'évaluer la position de p dans l'intervalle de confiance.

☑ Savoir faire : Savoir déterminer un intervalle de confiance:
On dispose d'une urne contenant un grand nombre de boules blanches et noires. La proportion de boules blanches contenues dans l'urne n'est pas connue. On réalise un tirage de 100 boules et on obtient 54 boules blanches. Estimer la proportion de boules blanches dans l’urne.

....
....
....
....
....

☑ Savoir faire : Savoir estimer une proportion inconnue par un intervalle de confiance:
Un institut de sondage interroge 1052 personnes entre les deux tours de l'élection présidentielle sur leur intention de vote. 614 déclarent avoir l'intention de voter pour un candidat A. En supposant que les votes seront conformes aux intentions, le candidat A a-t-il raison de croire qu'il sera élue ?

....
....
....
....
....
....
....

☑ Savoir faire : Savoir déterminer une taille d'échantillon suffisante pour obtenir une estimation :
Un constructeur automobile fait appel à un institut de sondage afin de mesurer le degré de satisfaction du service après-vente. L'institut souhaite estimer la proportion de clients satisfaits au niveau de confiance 0,95 avec une amplitude d'au plus 5 centièmes. Combien de personnes au minimum faut-il interroger ?

....
....
....
....
....
....
....
....

☑ *Amérique du nord 2013 :*

EXERCICE 2 **5 points**

Commun à tous les candidats

Dans cet exercice, les résultats seront donnés à 10^{-3} près.

1. Une étude interne à une grande banque a montré qu'on peut estimer que l'âge moyen d'un client demandant un crédit immobilier est une variable aléatoire, notée X, qui suit la loi normale de moyenne 40,5 et d'écart type 12.

a. Calculer la probabilité que le client demandeur d'un prêt soit d'un âge compris entre 30 et 35 ans.

b. Calculer la probabilité que le client n'ait pas demandé un prêt immobilier avant 55 ans.

2. Dans un slogan publicitaire, la banque affirme que 75% des demandes de prêts immobiliers sont acceptées.

Soit F la variable aléatoire qui, à tout échantillon de 1 000 demandes choisies au hasard et de façon indépendante, associe la fréquence de demandes de prêt immobilier acceptées.

a. Donner un intervalle de fluctuation asymptotique au seuil de 95% de la fréquence de prêts acceptés par la banque.

b. Dans une agence de cette banque, on a observé que, sur les 1 000 dernières demandes effectuées, 600 ont été acceptées. Énoncer une règle de décision permettant de valider ou non le slogan publicitaire de la banque, au niveau de confiance 95%.

c. Que peut-on penser du slogan publicitaire de la banque ?

………

………

………

………

………

………

………

………

………

………

………

………

………

☑ *Antilles–Guyane 19 juin 2013 :*

EXERCICE 4 **4 points**

Commun à tous les candidats

Les parties A et B sont indépendantes. Les résultats décimaux seront arrondis au millième pour tout l'exercice.

Partie A

La direction d'une société fabriquant des composants électroniques impose à ses deux sites de production de respecter les proportions ci-dessous en termes de contrat d'embauche du personnel :

— 80% de CDI (contrat à durée indéterminée)

— 20% de CDD (contrat à durée déterminée).

On donne la composition du personnel des deux sites dans le tableau suivant :

	CDI	CDD	Effectif total
Site de production A	315	106	421
Site de production B	52	16	68

1. Calculer le pourcentage de CDI sur chaque site de production.

2. Pour une proportion $p = 0,8$, déterminer les intervalles de fluctuation asymptotiques au seuil de 95% relatifs aux échantillons de taille n, pour $n = 421$ et pour $n = 68$.

3. Comment la direction de la société peut-elle interpréter les intervalles obtenus dans la question précédente ?

………

………

………

………

………

………

………

………

………

………

………

Partie B

Dans cette partie, on convient que l'on peut utiliser l'intervalle de fluctuation asymptotique lorsque $n > 30$, $np > 5$ et $n(1-p) > 5$, où p désigne la proportion dans une population, et n désigne la taille d'un échantillon de cette population.

La direction de cette même société tolère 7% de composants défectueux. Le responsable d'un site de production souhaite évaluer si sa chaîne de production respecte cette contrainte de 7%. Pour cela, il prélève un échantillon de composants électroniques.

1. S'il prélève un échantillon de 50 composants, peut-il utiliser l'intervalle de fluctuation asymptotique au seuil de 95%? Expliquer.
2. S'il prélève un échantillon de 100 composants, peut-il utiliser l'intervalle de fluctuation asymptotique au seuil de 95%? Expliquer.
3. Le responsable du site de production prélève un échantillon de taille 100, dans lequel 9 composants électroniques s'avèrent défectueux. Comment peut il interpréter ce résultat ?

☑ *Asie 2013 :*

EXERCICE 1 **4 points**

Commun à tous les candidats

Cet exercice est un QCM(questionnaire à choix multiples). Pour chacune des quatre questions, quatre réponses sont proposées ; une seule de ces réponses convient. Indiquer sur la copie le numéro de la question et la lettre de la réponse choisie sans justifier le choix effectué. Une bonne réponse rapporte 1 point. Une réponse fausse, une réponse multiple ou l'absence de réponse ne rapportent ni n'enlèvent aucun point.

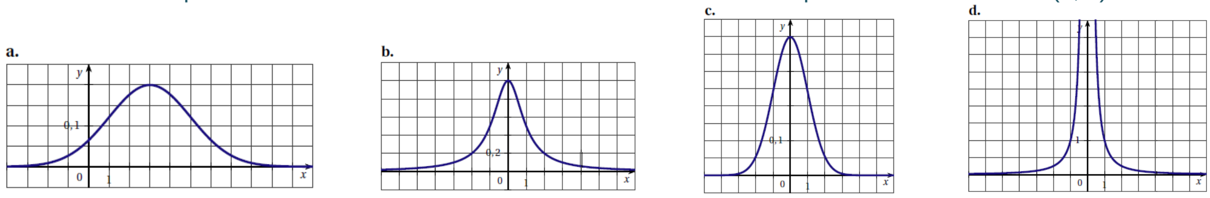
1. On choisit au hasard un réel de l'intervalle $[-2;5]$. Quelle est la probabilité que ce nombre appartienne à l'intervalle $[-1 ; 1]$?

a. $\frac{1}{5}$ **b.** $\frac{2}{7}$ **c.** $\frac{1}{2}$ **d.** 0,7

2. Soit X une variable aléatoire qui suit la loi normale de moyenne 3 et d'écart type 2.
Quelle est la valeur arrondie au centième de la probabilité $P(X \leq 1)$?

a. 0,16 **b.** 0,68 **c.** 0,95 **d.** 0,99

3. Quelle courbe représente la fonction de densité d'une variable aléatoire X qui suit la loi normale N (0, 1) ?



4. Lors d'un sondage avant une élection, on interroge 800 personnes (constituant un échantillon représentatif). 424 d'entre elles déclarent qu'elles voteront pour le candidat H.
Soit p la proportion d'électeurs de la population qui comptent voter pour H. Lequel des intervalles ci-dessous est un intervalle de confiance au niveau de confiance de 95% de la proportion p ?

a. $[0,46;0,60]$ **b.** $[0,48;0,58]$ **c.** $[0,49;0,57]$ **d.** $[0,51;0,55]$