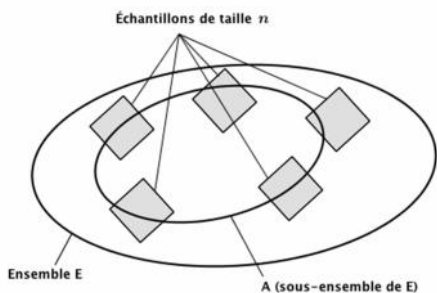


**Intervalle de fluctuation**

**Propriété** Supposons que la proportion d'un sous-ensemble A dans une population E est connue et vaut  $p$  (on peut aussi dire que, si l'on tire au hasard un individu dans la population E, la probabilité d'obtenir un individu dans A est égale  $p$ ). Considérons des échantillons de taille  $n$  dans cette population E. On peut établir que dans 95% de ces échantillons de taille  $n$ , la fréquence  $f$  des individus de A dans l'échantillon est dans



l'intervalle  $\left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$ .

Cet intervalle est appelé **intervalle de fluctuation au seuil de 95%**.

**Remarque:** Ce résultat n'est en fait valable que si  $0,2 \leq p \leq 0,8$  et si  $n \geq 25$  (autrement dit l'échantillon est de taille « suffisante »). L'objectif de cette activité est de comprendre et d'illustrer cette propriété.

**Première partie: expérience**

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 connaît la proportion  $p$  de boules rouges ; dans chaque biberon, cette proportion est égale à 0,3 (ou 30%).

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 "variabilité", cette "fluctuation", que nous allons observer et analyser. Les graphiques sont en annexe- page 4.

1. On décide de coder par 1 l'issue "une boule rouge sort" et 0 l'issue "une boule autre que rouge sort".

a) Noter les résultats obtenus après avoir réalisé 20 tirages de boules. On produit ainsi un échantillon de taille 20.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	

Effectif des boules rouges sur mon échantillon de taille 20 : .....

Fréquence des boules rouges sur mon échantillon de taille 20 :  $f = \dots\dots$

b) On renouvelle l'expérience avec un échantillon de taille 50.

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
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

Effectif des boules rouges sur mon échantillon de taille 50 : .....

Fréquence des boules rouges sur mon échantillon de taille 50 :  $f = \dots\dots$

c) On travaillera plutôt sur les fréquences observées que sur les effectifs observés. Pourquoi ?

2. À partir du récapitulatif des fréquences des boules rouges de la classe,

- a) Placer sur le graphique 1, les points ayant pour abscisse 1, 2, 3, ... et pour ordonnée la fréquence  $f$  des boules rouges obtenue par chaque élève de la classe sur son échantillon de taille 20.
- b) Placer sur le graphique 3, les points ayant pour coordonnées  $(n ; f)$  : le premier "lot" de points correspond aux points ayant 20 comme abscisse commune et pour ordonnée la fréquence  $f$  des boules rouges obtenues par chaque élève de la classe. De même, le second "lot" correspond aux points d'abscisse commune 50.

**Deuxième partie : simulation (en salle informatique)**

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		

- 1. a) Expliquer la formule entrée dans les cellules B2 et G2.
- b) Jusqu'à quelle ligne faut-il recopier les formules des cellules B2 et G2 ?
- c) Quelles formules faudra-t-il entrer dans les cellules D2, D3, I2 et I3 ?
  
- 2. a) Ouvrir le fichier « simulation- biberon », puis le compléter. Appeler le professeur pour valider la feuille de calcul.
- b) En appuyant sur la touche F9 ou « ctrl flèche F9 », compléter le tableau suivant avec 24 fréquences de boules rouges sur les échantillons de taille 100, puis de taille 500.

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

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

3. a) Sur le graphique 2, placer les points de coordonnées  $(i ; f)$  en rouge pour les échantillons de taille 100 et en vert pour les échantillons de taille 500

b) Compléter le graphique 3.

**Troisième partie : description des séries statistiques.**

1.a) À l'aide du récapitulatif des fréquences des boules rouges et en utilisant un outil adapté, déterminer les indicateurs demandés pour les quatre séries statistiques obtenues :

	Série des fréquences observées sur des échantillons de taille 20	Série des fréquences observées sur des échantillons de taille 50	Série des fréquences observées sur des échantillons de taille 100	Série des fréquences observées sur des échantillons de taille 500
Minimum				
1 <sup>er</sup> quartile				
Médiane				
3 <sup>e</sup> quartile				
Maximum				

b) Construire (éventuellement sous GeoGebra) le diagramme contenant les quatre boîtes à moustaches de ces séries.

2. Que peut-on tirer des différentes illustrations graphiques que l'on a faites tout au long de l'activité ? Quel est l'effet de la taille de l'échantillon ?

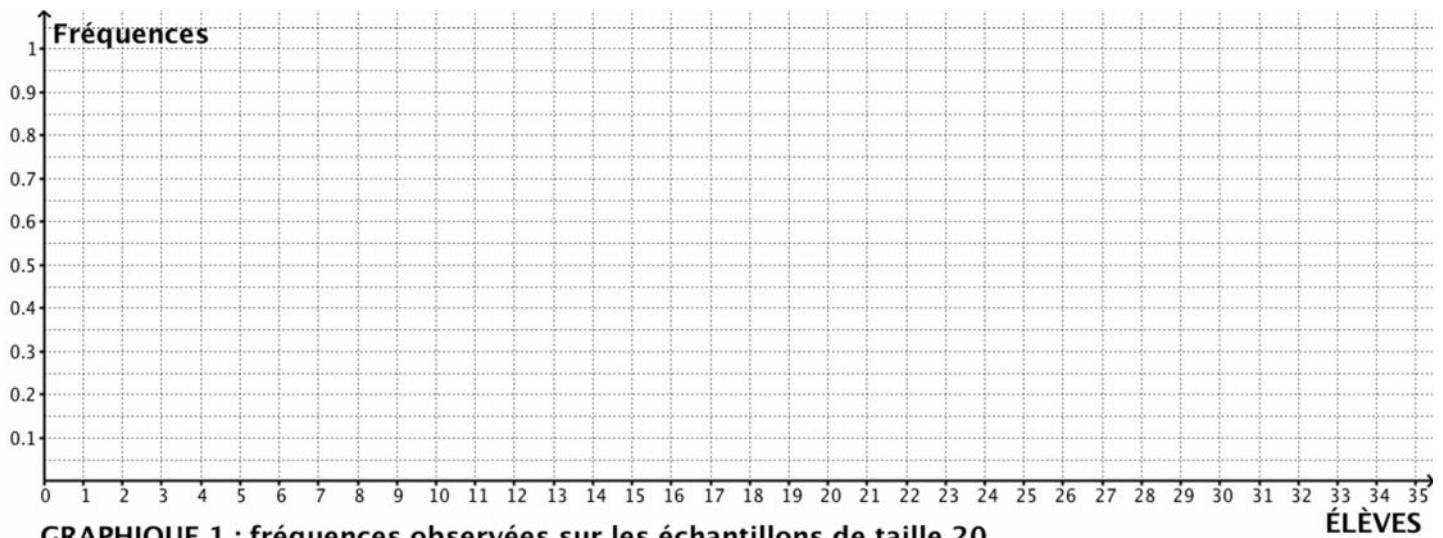
3. a) Sur le graphique 3, placer en rouge, les points A et B de coordonnées respectives  $\left(n; 0,3 + \frac{1}{\sqrt{n}}\right)$  et

$\left(n; 0,3 - \frac{1}{\sqrt{n}}\right)$  avec  $n = 20$

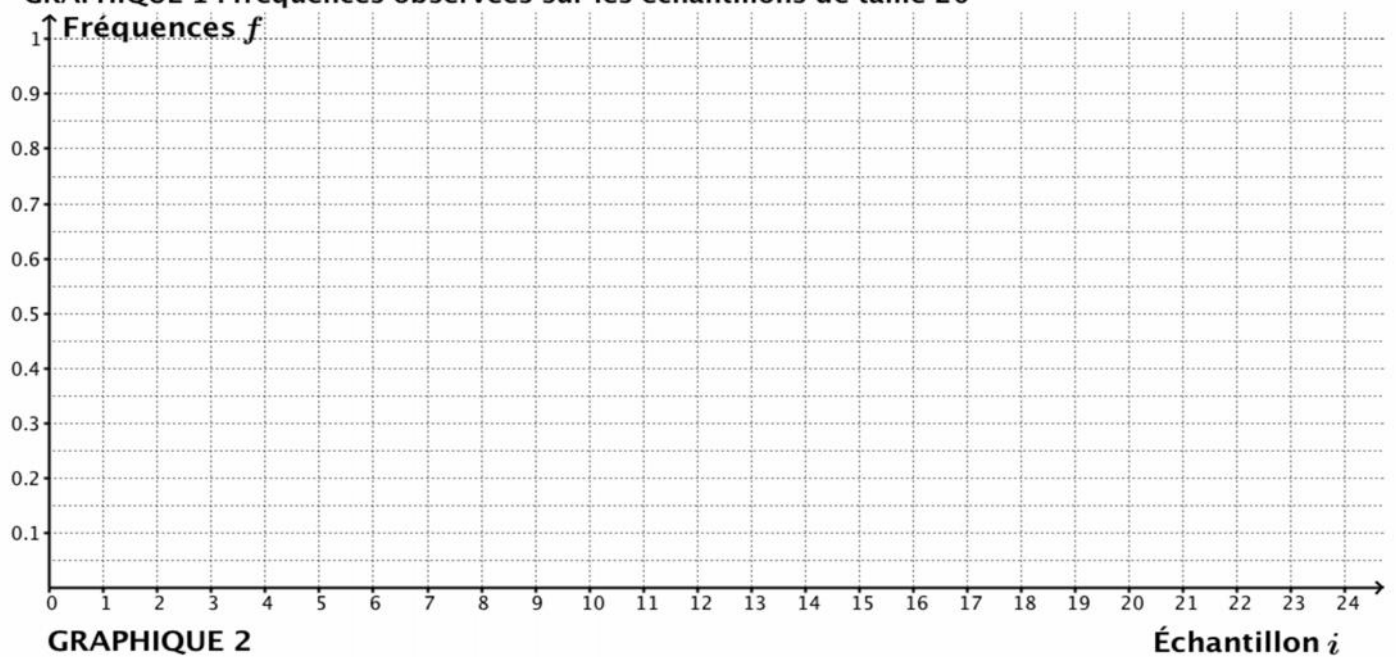
b) Déterminer le pourcentage de points appartenant au segment [AB]

4. Reprendre les questions 3.a) et 3.b) pour  $n = 50$  (on notera C et D les deux points) ;  $n = 100$  (on notera E et F les deux points) et  $n = 500$  (On notera G et H les deux points)

5. Conclure par rapport à la propriété donnée en début d'activité.



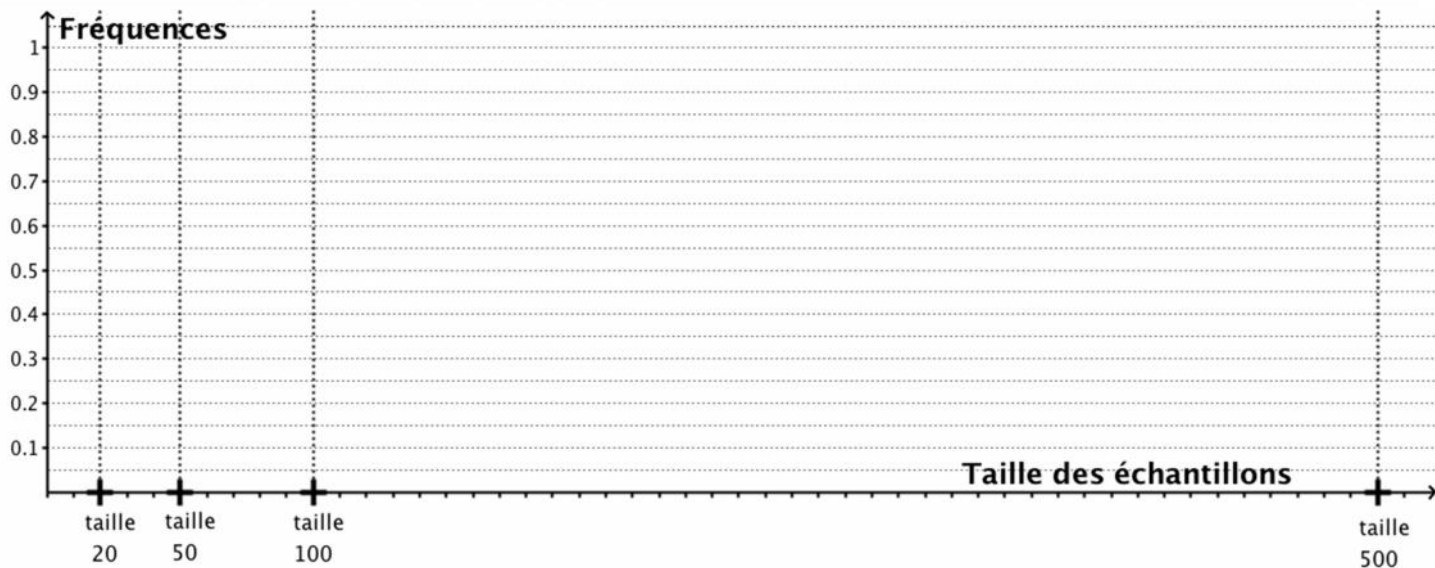
GRAPHIQUE 1 : fréquences observées sur les échantillons de taille 20



GRAPHIQUE 2

En rouge : fréquences observées sur les échantillons de taille 100

En vert : fréquences observées sur les échantillons de taille 500



GRAPHIQUE 3