

Un biberon comme outil de simulation au lycée

Nelly LASSALLE et Annette CORPART (IREM de Clermont-Ferrand)

I – Découvrir la fluctuation d'échantillonnage (programme de seconde)

On considère une population constituée de 30% de boules rouges et de 70% de boules blanches.

Le nombre de boules rouges obtenu dans un échantillon de taille n prélevé dans la population est variable. C'est une variable aléatoire X_n .

L'objectif est de comprendre et d'illustrer la variabilité de la fréquence $F_n = \frac{X_n}{n}$ d'apparition des boules rouges dans l'échantillon.

1° Expérience et recueil des données.

Chaque élève reçoit une bouteille opaque munie d'une tétine et **connaît la population** à l'intérieur (soit 3 rouges et 7 blanches, soit 6 rouges et 14 blanches). Le bouchon est percé d'un trou suffisamment large pour laisser échapper une boule dans la tétine si on retourne la bouteille : on note 1 à chaque fois que la boule est rouge et 0 sinon.

On va prélever des échantillons aléatoires de taille n dans la population. Pour cela, on retourne n fois la bouteille, sans oublier après chaque « retournement » de remettre la boule dans la bouteille et de bien secouer.

On rappelle qu'un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience, l'indépendance étant assurée dans cette expérience par la remise de la boule.

- Réaliser alors 20 tirages de boules pour générer un échantillon A de taille 20 et noter les résultats (0 ou 1) dans la fiche individuelle (page 4).

De la même manière, noter les résultats d'un échantillon B obtenu avec 50 tirages.

Calculer ensuite le cumul des 1 pour chacun des deux échantillons. On a ainsi le nombre X_n de boules rouges obtenu dans chaque échantillon.

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

Calculer (mentalement) les fréquences de boules rouges obtenues dans les deux échantillons.

- Pour obtenir des échantillons de grande taille, on va utiliser un tableur.

Si on tape $\text{Ent}(\text{Rand} + 0,30)$, la calculatrice donne 1 dans 30% des cas et 0 dans 70% des cas. Pourquoi ?

Simuler sur un tableur des échantillons C et D de tailles respectives 100 et 500. Pour chacun de ces deux échantillons, compter le nombre de boules rouges obtenu et calculer la fréquence correspondante.

- La fiche récapitulative de la classe pour les quatre échantillons se remplit manuellement (page 5). L'enseignant peut ensuite en donner une photocopie à chaque élève.

2° Description des séries statistiques.

A l'aide du tableur et du logiciel Géogébra, on va illustrer les fréquences $F_n = \frac{x_n}{n}$ des boules rouges par des graphiques statistiques.

- Dans une nouvelle page du tableur, en supposant que la classe comporte N élèves, remplir les N premières cellules de la première colonne avec la taille des échantillons A (soit 20) et les N premières cellules de la deuxième colonne avec les fréquences relevées page 5. Procéder de la même manière dans les colonnes suivantes pour les échantillons B, C et D.

Le tableur peut calculer des paramètres statistiques pour les quatre séries : moyenne, médiane, quartiles...

- Construire sur un même graphique les quatre nuages de points associés aux quatre séries. Choisir l'option *graphique à nuage de points non reliés* et faire apparaître pour chaque série les points de coordonnées (n, f) . Voir annexe 1 page 6.

- Copier les données statistiques relevées précédemment puis les coller dans le tableur de Géogébra (tableur à cocher dans l'option *Affichage*). Construire sur un même graphique les quatre diagrammes en boîte associés aux quatre séries. Dans *Commande*, choisir l'option *BoiteMoustaches* et indiquer comme syntaxe dans les crochets : valeur de l'ordonnée pour l'axe de la boîte, demi-hauteur de la boîte, plage des données statistiques.

Exemple de l'annexe 2 page 7 : BoiteMoustaches [1, 0.5, B1:B30] pour l'échantillon A.

On peut ajuster les proportions du graphique avec l'option *Recadrer*, après un clic droit sur la souris.

L'enseignant pourra, à la suite de ce travail, demander à chaque élève d'imprimer le graphique « nuages » et le graphique « boîtes ».

Remarque :

On peut également visualiser le graphique « boîtes » sur une calculatrice (voir les instructions ci-dessous) mais la copie sur papier est moins simple.

- **CASIO** : menu stat → saisir les données de la série des échantillons A, dans List 1.
GRPH → ► → SET → GPH1 → G-Type: ► Box → XList: List1 → Freq: 1 → (EXE) → GPH1 → (s'il y a eu un autre graphique avant, faire (Sketch) → Cls). Penser à régler la fenêtre (V-Window) pour obtenir un graphique lisible.
Pour noter les valeurs des quartiles, de la moyenne et de l'écart-type faire : (Trace) → ► ► ► ► ► .
Faire de même avec les séries des échantillons B, C et D, il suffit de changer le numéro de la liste, sans oublier de faire (Sketch) → Cls après chaque graphique.
- **TEXAS** : saisir les données de la série des échantillons A dans L1.
(STAT PLOT) → Graph1 ← → Aff ← → Type: (picto.boite) ← → ListeX: L1 → Effectifs: 1 → (GRAPH). Penser à régler la fenêtre (WINDOW) pour obtenir un graphique lisible.
Pour noter les valeurs des quartiles faire : (Trace) → ► ► ► ► ► .
Pour moyenne et écart-type faire : STAT ► CALC ► 1-Var Stats ← ←
Faire de même avec les séries des échantillons B, C et D, il suffit de changer le numéro de la liste.

3° **Analyser et comparer les quatre distributions observées de la variable** $F_n = \frac{X_n}{n}$.

Afin de comparer plusieurs échantillons, la juxtaposition judicieuse des graphiques permet d'observer la dispersion et la position des quatre séries et d'émettre des hypothèses sur les facteurs qui en sont à l'origine.

Que peut-on tirer des illustrations graphiques que l'on vient de faire ? Quel est l'effet de la taille d'échantillon ?

- Tracer sur chacun des nuages de points le plus petit intervalle centré sur 0,30 contenant au moins 95 % des fréquences observées.

En classe de seconde, le programme définit l'intervalle de fluctuation d'une proportion p d'un caractère dans la population comme étant l'intervalle centré autour de p où se situe, avec une probabilité au moins égale à 0,95, la fréquence observée f dans un échantillon de taille n . Pour des proportions p du caractère comprises entre 0,2 et 0,8, et pour des échantillons de taille $n \geq 25$, 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.

- Déterminer l'intervalle de fluctuation au niveau 95% de la proportion de boules rouges dans un échantillon de taille 50, 100 ou 500.

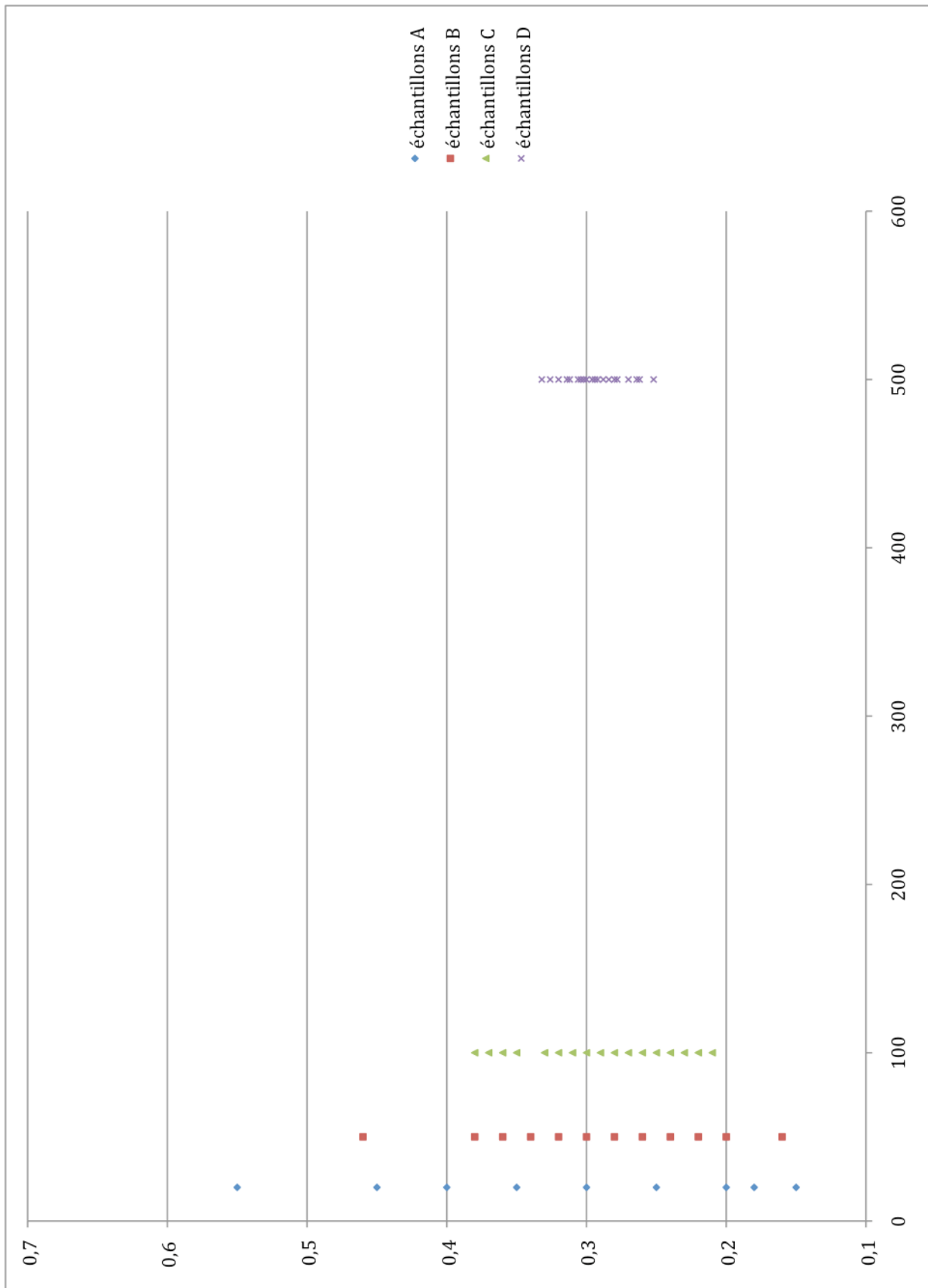
Comparer avec les intervalles tracés précédemment.

Conclusion : dans 95% des cas, on n'observe pas de différence significative entre p et les valeurs de $F_n = \frac{X_n}{n}$.

Fiche individuelle (I – 1°)

NOM Prénom :					Effectifs de boules rouges
				Echantillon B	tirage 1
				Echantillon B	tirage 2
		Effectifs de boules rouges		Echantillon B	tirage 3
				Echantillon B	tirage 4
Echantillon A	tirage 1			Echantillon B	tirage 5
Echantillon A	tirage 2			Echantillon B	tirage 6
Echantillon A	tirage 3			Echantillon B	tirage 7
Echantillon A	tirage 4			Echantillon B	tirage 8
Echantillon A	tirage 5			Echantillon B	tirage 9
Echantillon A	tirage 6			Echantillon B	tirage 10
Echantillon A	tirage 7			Echantillon B	tirage 11
Echantillon A	tirage 8			Echantillon B	tirage 12
Echantillon A	tirage 9			Echantillon B	tirage 13
Echantillon A	tirage 10			Echantillon B	tirage 14
Echantillon A	tirage 11			Echantillon B	tirage 15
Echantillon A	tirage 12			Echantillon B	tirage 16
Echantillon A	tirage 13			Echantillon B	tirage 17
Echantillon A	tirage 14			Echantillon B	tirage 18
Echantillon A	tirage 15			Echantillon B	tirage 19
Echantillon A	tirage 16			Echantillon B	tirage 20
Echantillon A	tirage 17			Echantillon B	tirage 21
Echantillon A	tirage 18			Echantillon B	tirage 22
Echantillon A	tirage 19			Echantillon B	tirage 23
Echantillon A	tirage 20			Echantillon B	tirage 24
	Somme			Echantillon B	tirage 25
				Echantillon B	tirage 26
				Echantillon B	tirage 27
				Echantillon B	tirage 28
				Echantillon B	tirage 29
				Echantillon B	tirage 30
				Echantillon B	tirage 31
				Echantillon B	tirage 32
				Echantillon B	tirage 33
				Echantillon B	tirage 34
				Echantillon B	tirage 35
				Echantillon B	tirage 36
				Echantillon B	tirage 37
				Echantillon B	tirage 38
				Echantillon B	tirage 39
				Echantillon B	tirage 40
				Echantillon B	tirage 41
				Echantillon B	tirage 42
				Echantillon B	tirage 43
				Echantillon B	tirage 44
				Echantillon B	tirage 45
				Echantillon B	tirage 46
				Echantillon B	tirage 47
				Echantillon B	tirage 48
				Echantillon B	tirage 49
				Echantillon B	tirage 50
					Somme

Annexe 1



Annexe 2

