Un biberon comme outil de simulation au lycée

Annette Corpart et Nelly Lassalle(*)

Depuis 2010, à la demande de l'inspection régionale, nous intervenons dans la formation continue des enseignants pour la partie probabilités/statistiques des nouveaux programmes de lycée. Cette formation a eu lieu trois ans de suite dans les quatre départements de l'académie de Clermont-Ferrand et a concerné environ 250 professeurs chaque année. Cette intervention se fait sous la forme d'un exposé théorique d'une part et d'un atelier d'autre part.

C'est dans le cadre de cet atelier que nous avons imaginé une activité découpée en trois parties, chacune des parties introduisant une notion différente : fluctuation d'échantillonnage en seconde, prise de décision en première et estimation d'une proportion en terminale. Cette activité peut être distribuée telle quelle aux élèves, les parties du texte en italique et les annexes étant destinées à l'usage du professeur.

Nous sommes parties d'une situation concrète rencontrée en lycée agricole (le LEGTA Louis Giraud de Serres dans le Vaucluse). Pour effectuer les tirages, le collègue utilise des bouteilles contenant des billes et peintes en noir jusqu'au goulot pour laisser visible la couleur de la bille tirée. Nous avons amélioré le procédé en utilisant des bouteilles opaques de lait, dans lesquelles nous découpons le bouchon pour introduire une tétine transparente (d'où notre appellation de « biberon »). À l'intérieur de la bouteille, nous mettons des boules de cotillon (moins bruyantes que des billes) et ainsi le contenu de la bouteille demeure caché pendant toute l'activité.

Nous avons testé cette activité dans nos classes de seconde. Les parties première et terminale ont été testées (en les adaptant) dans les classes de BTS industriels où les statistiques inférentielles sont enseignées. Dans toutes ces classes, nous avons eu une excellente réceptivité de la part des lycéens comme des étudiants.

Les enseignants qui ont assisté aux journées de formation ont également été mis en « situation élève ». Nous sommes persuadées que cette activité permet une bonne appréhension des notions difficiles de statistiques inférentielles.

Rappelons enfin que Guy Brousseau a travaillé en 1973 et 1974 dans quatre classes de CM2 en utilisant des bouteilles en plastique dont seul le goulot était transparent et contenant des billes de deux couleurs dont la répartition était inconnue des élèves. Il en a rendu compte dans plusieurs de ses publications.

(*) 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 **l'enseignant donne la proportion de boules rouges à l'intérieur**, soit 30%. Le bouchon est percé d'un trou suffisamment large pour laisser descendre 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 et l'agitation de la bouteille.

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. 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 **= Ent(Alea() + 0,30)** (avec le tableur de géogébra, on tape **floor(random() + 0.3))**, la calculatrice donne 1 avec une probabilité de 30 % et 0 avec une probabilité de 70 %.

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 (voir annexe 1). L'enseignant peut ensuite en donner une photocopie à chaque élève.

2° Description des séries statistiques.

À l'aide du logiciel Géogébra (version 4), on va illustrer les fréquences $F_n = \frac{X_n}{n}$ des boules rouges par des graphiques statistiques.

Dans une page du tableur (à cocher dans l'option **Affichage***), 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. 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 papier millimétré les quatre nuages de points de coordonnées (n, f) *associés aux quatre séries (voir annexe 2).*

Ce graphique peut être obtenu sur le tableur en sélectionnant les colonnes deux par deux et en choisissant l'option statistiques à deux variables. Par un clic droit sur la fenêtre, on peut exporter chaque nuage obtenu vers le graphique (voir annexe 3). L'inconvénient du tableur est que les points de mêmes coordonnées se superposent.

Revenir sur la page du tableur. Construire sur un même graphique les quatre diagrammes en boîte associés aux quatre séries. Dans **Saisir***, écrire* **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 4 : **BoiteMoustaches [1, 0.5, B2:B18]** pour l'échantillon A.

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

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 $\mathbf{F}_n = \dfrac{\mathbf{X}_n}{n}$

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 étudié dans la population comme étant le plus petit 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 $\left[p - \dfrac{1}{\sqrt{n}} ; p + \dfrac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95.

Avec la formule précédente, 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 cette activité, la fréquence *f* observée est en dehors de l'intervalle de fluctuation dans environ 5% des cas. Autrement dit, dans 95 % des cas, on n'observe pas de différence significative entre *p* et les valeurs de** $\mathbf{F}_n = \dfrac{\mathbf{X}_n}{n}$**.**

II – Estimer une proportion (programme de terminale)

Une entreprise d'agronomie produit en grande quantité de la semence pour gazon, qui est un mélange de deux types de graines :

- de l'Agropyrum Repens (chiendent, noté AR), dans une proportion p. Cette graine est adaptée aux conditions sèches et au piétinement mais n'offre pas une couverture très esthétique.
- du Lolium Perenne (ray-grass en anglais), dans une proportion $1 - p$.

Le nombre de graines d'AR obtenu dans un échantillon aléatoire de taille n est variable. C'est une variable aléatoire notée X_n.

L'objectif est d'estimer p à partir de la fréquence $F_n = \dfrac{X_n}{n}$ d'apparition des graines d'AR dans un échantillon.

1° Expérience et recueil des données.

On va prélever un échantillon de n graines, successivement et avec remise, dans la production de l'entreprise contenant **une proportion p fixée mais inconnue de graines d'AR**. Cette production sera représentée par une bouteille opaque munie d'une tétine et contenant des boules rouges et des boules d'une autre couleur. À chaque tirage, on notera 1 à chaque fois que la couleur de la « graine » obtenue est rouge (c'est une graine d'AR) et 0 sinon.

Remarquons que dans la réalité, on prélèverait n graines au hasard mais sans remise ; l'échantillon serait pourtant considéré comme aléatoire car obtenu parmi un nombre considérable de graines. D'autre part, la reconnaissance effective des graines se fait selon leur morphologie à l'œil nu ou à la loupe (certaines filières de BTSA ont les graines de 100 espèces à reconnaître en utilisant des flores de reconnaissance).

Réaliser alors 50 tirages de graines pour générer un échantillon de taille 50 et noter les résultats (0 ou 1) dans la fiche individuelle. Calculer le cumul des 1 puis la fréquence f des graines d'AR pour votre échantillon.

En comparant avec les résultats obtenus par les autres élèves, est-il raisonnable de prendre la valeur f de votre observation comme estimation de la proportion p ? Pourquoi ?

2° Estimer p par une fourchette.

On sait depuis la classe de seconde que : **pour une proportion p du caractère comprise entre 0,2 et 0,8 et pour des échantillons de taille $n \geq 25$, f appartient à l'intervalle de fluctuation $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95.**

Construire un algorithme permettant d'afficher les bornes des intervalles de fluctuation pour des valeurs de p variant de 0,2 à 0,8 avec un pas de 0,05 puis représenter ces intervalles sur papier millimétré : à chaque valeur de p en abscisse, on dessinera verticalement l'intervalle associé (voir annexe 5).

On peut remarquer que les bornes de tous les intervalles de fluctuation que l'on pourrait fabriquer se situent sur les droites d'équations $y = x - \frac{1}{\sqrt{50}}$ et $y = x + \frac{1}{\sqrt{50}}$.

Tracer ces droites.

Réciproquement, en plaçant en ordonnée la valeur de f obtenue au 1° avec un échantillon de 50 graines, on peut déterminer graphiquement les intervalles de fluctuation contenant f avec une probabilité d'au moins 0,95. Pour cela, on cherche les abscisses a et b des points d'intersection des deux droites précédentes avec la droite d'équation $y = f$ (*voir annexe 6*).

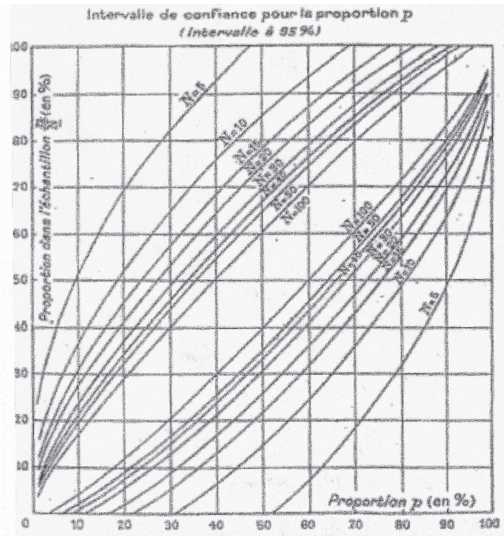
[a ; b] est appelé un intervalle de confiance de p au niveau 95 %. Ceci signifie que p est en dehors de cet intervalle pour seulement 5 % des échantillons.

Donner votre estimation de la proportion de graines d'AR dans la production de semence par un intervalle de confiance au niveau 95%.

3° Remarques.

• Selon l'échantillon de 50 graines tiré par chaque élève, la valeur de f obtenue peut être différente. On ne peut donc pas parler de **l'**intervalle de confiance de p mais seulement d'**un** intervalle de confiance de p.

• Avec le travail sur la loi normale, on montre que l'intervalle de fluctuation de p au seuil 95% est de la forme $\left[p-1{,}96\dfrac{\sqrt{p(p-1)}}{\sqrt{n}}\,;\,p+1{,}96\dfrac{\sqrt{p(p-1)}}{\sqrt{n}}\right]$ dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$. On admettra que si l'on dessine les intervalles de fluctuation pour des valeurs de p variant de 0 à 1, leurs bornes se situent sur des arcs d'ellipses (*voir annexe 6*).



Le professeur pourra dévoiler la composition des biberons et donner la valeur de la proportion de graines d'AR (par exemple p = 0,25) même si dans l'entreprise cette opération est impossible à réaliser.
Les élèves peuvent ainsi s'apercevoir que leur propre intervalle de confiance ne contient pas forcément la valeur de p. Ils calculent alors le pourcentage de « bons » intervalles dans la classe et s'aperçoivent que ce résultat est proche de 95%.

III – Prendre une décision dans un contrôle de qualité (programme de première)

Dans le service technique d'une commune, vous êtes chargé de réceptionner la commande d'un lot de semence (par exemple une tonne) pour engazonner les espaces verts. Le cahier des charges de la commande précise que **le mélange doit contenir**

30% de graines d'Agropyrum Repens (chiendent, noté AR). Cette plante n'offre pas une couverture très esthétique mais elle est bien adaptée à la sécheresse et au piétinement.
La procédure de contrôle statistique, appelé plan de contrôle, fixe :

- la procédure d'échantillonnage en indiquant **la taille *n* de l'échantillon aléatoire** de graines à prélever dans le lot et les modalités de prélèvement. On compte le nombre de graines d'AR dans cet échantillon.
- **le critère de rejet du lot** : c'est le nombre de graines d'AR à partir ou en dessous duquel le lot doit être rejeté comme non conforme à la commande.

Deux erreurs peuvent survenir dans votre prise de décision : vous pouvez vous tromper en rejetant un lot conforme (risque fournisseur) ou en acceptant un lot non conforme (risque client). Ces risques proviennent de la variabilité des résultats issus d'une procédure d'échantillonnage : en échantillonnant deux fois de la même façon, on peut obtenir des résultats différents (voir programme de seconde).

Dans un premier temps, on réalise l'expérience pour recueillir les données. Dans un deuxième temps, on travaille sur un modèle probabiliste pour comprendre comment obtenir les critères de rejet pour un risque d'erreur fixé.

1° Protocole expérimental et recueil des données.

Comme un contrôle réel est impossible en classe, un lot de semence sera représenté par une bouteille opaque munie d'une tétine et contenant des boules rouges et des boules d'une autre couleur, en proportion inconnue.
Chaque élève expérimente avec une bouteille numérotée, assimilée au lot de semence.
À chaque tirage, on notera 1 si la couleur de la « graine » obtenue est rouge (c'est une graine d'AR) et 0 sinon.
Remarquons que pour un contrôle réel, on prélèverait *n* graines au hasard mais sans remise ; l'échantillon serait pourtant considéré comme aléatoire car obtenu parmi un nombre considérable de graines.

Réaliser 20 tirages de graines pour générer un échantillon A de taille 20 et noter les résultats (0 ou 1) dans votre fiche individuelle. De la même manière, noter les résultats d'un échantillon B obtenu avec 50 tirages.
Calculer ensuite le nombre X *de graines d'AR obtenues dans chacun des deux échantillons.*

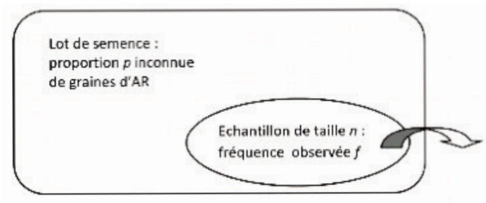
Le tableau récapitulatif de la classe se remplit manuellement et est photocopié ensuite par l'enseignant.

2° Retour sur le contrôle de qualité.

Vous devez prendre une décision concernant l'acceptation ou le refus du lot de semence livré.
Qu'est ce qu'un lot conforme au cahier des charges ?
Quels lots vous semblent non conformes à la commande au vu des échantillons A ? Des échantillons B ?

Les enseignants qui ont testé cette activité dans les journées de formation ont eu des réactions variées : certains ont eu tendance à rejeter beaucoup de lots, d'autres pas. Ils sont alors tombés d'accord sur la nécessité d'une règle de décision.

3° Construction du critère de rejet à l'aide des probabilités.

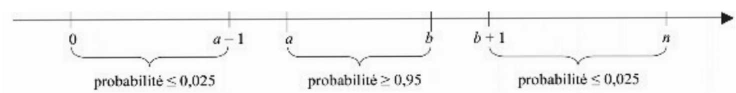


Prenons comme **hypothèse** que **la proportion *p* est de 30%** car c'est celle qui est prévue par le cahier des charges. On rejettera cette hypothèse lorsqu'on observera une différence significative entre la valeur supposée de p et la valeur observée f (donc si f est trop éloignée de p).

On choisit un seuil de décision α : on souhaite que la probabilité de rejeter l'hypothèse alors qu'elle est vraie, soit au maximum de 5% lorsqu'on choisit $\alpha = 0{,}05$.

On rappelle que X est la variable aléatoire mesurant le nombre de graines d'AR observées dans un échantillon de taille n. *Quelle est la loi suivie par* X *? Quels sont ses paramètres ?*

On cherche à partager l'intervalle $[0, n]$, où X prend ses valeurs, en trois intervalles $[0\ ;\ a-1]$, $[a\ ;\ b]$ et $[b+1\ ;\ n]$ de sorte que X prenne ses valeurs dans chacun des intervalles extrêmes avec une probabilité cumulée d'au plus 0,025, ce qui est réalisé si $P(X \leq a-1) \leq 0{,}025$ et $P(X \geq b+1) \leq 0{,}025$.



Comme la loi de X est discrète, on peut remarquer que :

$$P(X \leq a-1) \leq 0{,}025 \Leftrightarrow P(X < a) \leq 0{,}025$$

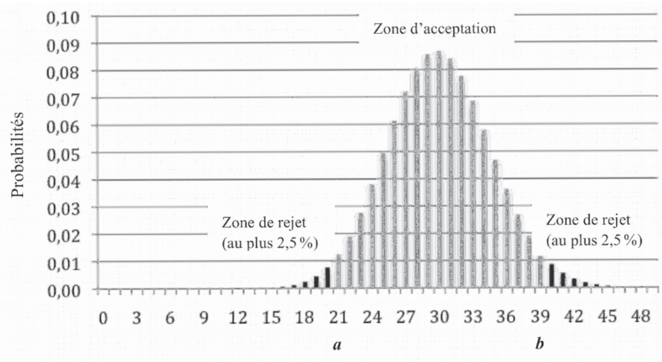
et

$$P(X \geq b+1) \leq 0{,}025 \Leftrightarrow P(X > b \ \leq 0{,}025.$$

Il suffit donc de déterminer :

- le plus petit entier a tel que $P\left(X \leq a\right) > 0{,}025 \Leftrightarrow \sum_{k=0}^{a} \binom{n}{k} \times 0{,}3^k \times 0{,}7^{n-k} > 0{,}025$,,
- le plus petit entier b tel que $P\left(X \leq b\right) \geq 0{,}975 \Leftrightarrow \sum_{k=0}^{b} \binom{n}{k} \times 0{,}3^k \times 0{,}7^{n-k} \geq 0{,}975$.

Exemple avec la loi binomiale $\mathcal{B}(100 ; 0{,}3)$:



La règle de décision est la suivante : **si la fréquence observée *f* de graines d'AR appartient à l'intervalle $\left[\frac{a}{n} ; \frac{b}{n}\right]$, on considère que l'hypothèse selon laquelle le lot de semence comporte 30 % d'AR n'est pas remise en question et on l'accepte ; sinon, on rejette cette hypothèse en sachant qu'on a 5 % de risque de faire une erreur.**

Les entiers *a* et *b* se déterminent, soit à l'aide d'un algorithme, soit avec un tableur.

Études l'algorithme suivant :

<u>Variables</u>
n, p, α, k, S
<u>Entrée</u>
Saisir α
Saisir n
Saisir p
k prend la valeur 0
S prend la valeur 0
<u>Traitement</u>
Tant que $S \leq \frac{\alpha}{2}$ faire
  S prend la valeur $S + \binom{n}{k} \times p^k \times (1-p)^{n-k}$
  k prend la valeur $k + 1$
Fin Tant que
<u>Sortie</u>
Afficher $k - 1$

Identifier les différentes variables de cet algorithme.

L'appliquer pour $\alpha = 0{,}05$, $p = 0{,}30$ *et* $n = 20$ (*ou* $n = 50$). *Quelle valeur de* α

obtient-on ?

Modifier l'algorithme précédent pour obtenir également la valeur de b.
Programmer l'algorithme sur votre calculatrice.

Avec un tableur (ordinateur ou calculatrice), on peut également obtenir les valeurs de a et b en observant une table de valeurs prises par $P(X \leq k)$.

Avec un ordinateur (*voir annexe 7*), on procède de la manière suivante, décrite pour le cas $n = 50$ et $p = 0{,}30$:

- Dans la colonne A, on écrit les valeurs de k de 0 à 50. (par exemple 0 en A3, 1 en A4, …)
- En entrant la formule **= LOI.BINOMIALE (A3; 50; 0,30; FAUX)** dans la cellule B3, on calcule la probabilité P(X = A3), c'est-à-dire P(X = 0) dans notre exemple.
- En entrant la formule **= LOI.BINOMIALE (A3; 50; 0,30; VRAI)** dans la cellule C3, on calcule les probabilités cumulées P(X ≤ A3), c'est-à-dire P(X ≤ 0) dans notre exemple.
- On recopie ces deux formules en colonne pour les valeurs de k de 0 à 50.
  Avec « Openofficecalc », il faut écrire 0 à la place de FAUX et 1 à la place de VRAI.

Avec une calculatrice, on utilise l'éditeur de fonctions, en entrant la formule **BinomialCD(X,50,0.30)** pour Casio ou **binomFRép(50,0.30,X)** pour Texas.

La valeur de a est le plus grand entier k tel que $P(X \leq k) > 0{,}025$; celle de b est le plus petit entier k tel que $P(X \leq k) \geq 0{,}975$.

En utilisant une des deux méthodes, compléter le tableau suivant :

Taille d'échantillon	Valeur de a	Valeur de b	Intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$
n = 20			
n = 50			

Quels sont les lots qui sont refusés ?

Remarque : Pour n assez grand et p ni trop petit ni trop grand, on observe que l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$ est sensiblement le même que l'intervalle de fluctuation $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ étudié en seconde. *Le vérifier.*

L'intérêt de l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$, calculé à partir de la loi binomiale, **est de convenir pour toutes les valeurs des paramètres n et p** (en particulier pour de petits échantillons) alors que l'intervalle $\left[p-\dfrac{1}{\sqrt{n}}\,;p+\dfrac{1}{\sqrt{n}}\right]$ n'est adapté que sous certaines conditions de n et de p.

4° Analyse des erreurs.

On dévoile la composition des bouteilles numérotées qui contiennent chacune vingt boules. Certaines ont une proportion de boules rouges égale à 30%, d'autres une proportion de boules rouges inférieure ou supérieure à 30%.

Relever cette proportion pour chacune des bouteilles sur le tableau récapitulatif.

Noter les cas où l'on s'est trompé.

On fera bien la différence entre les deux types d'erreur :

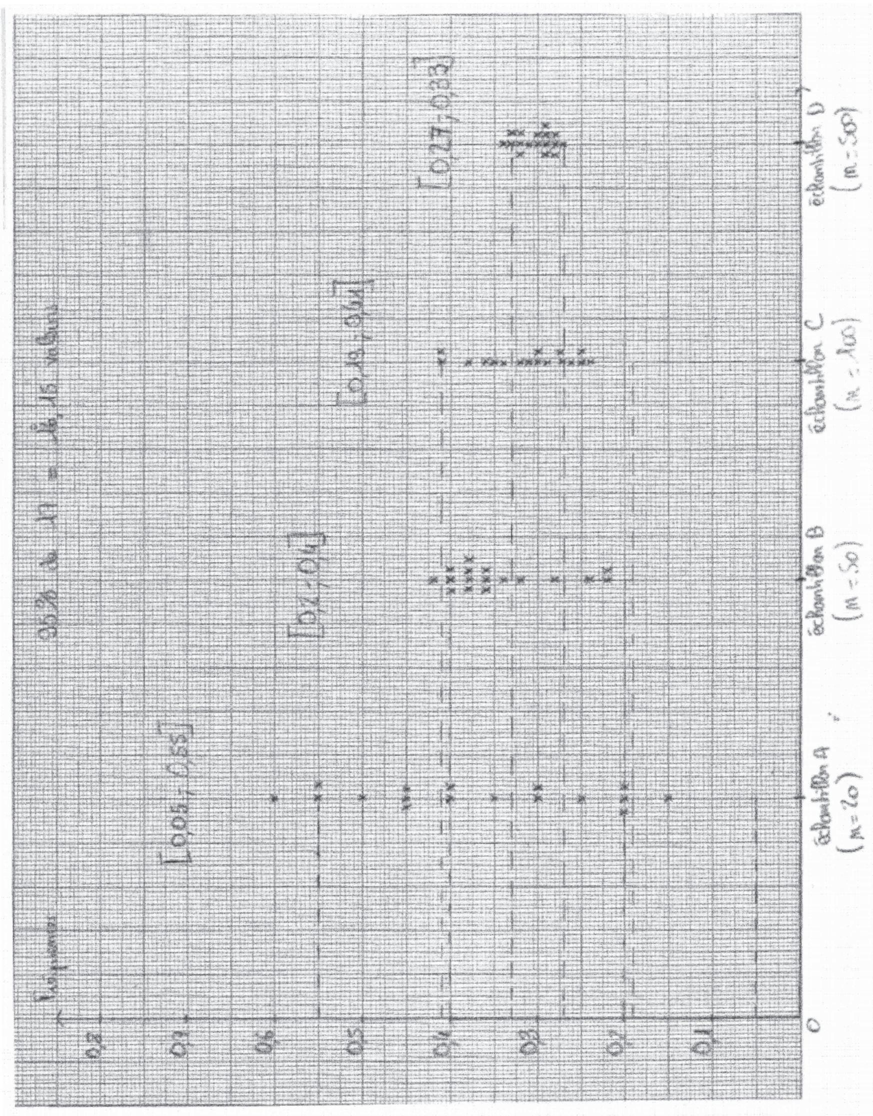
- on a rejeté un lot conforme (surligner en vert) : c'est dommage pour le fournisseur.
- on a accepté un lot non conforme (surligner en rouge) : c'est dommage pour notre gazon !

Annexe 1

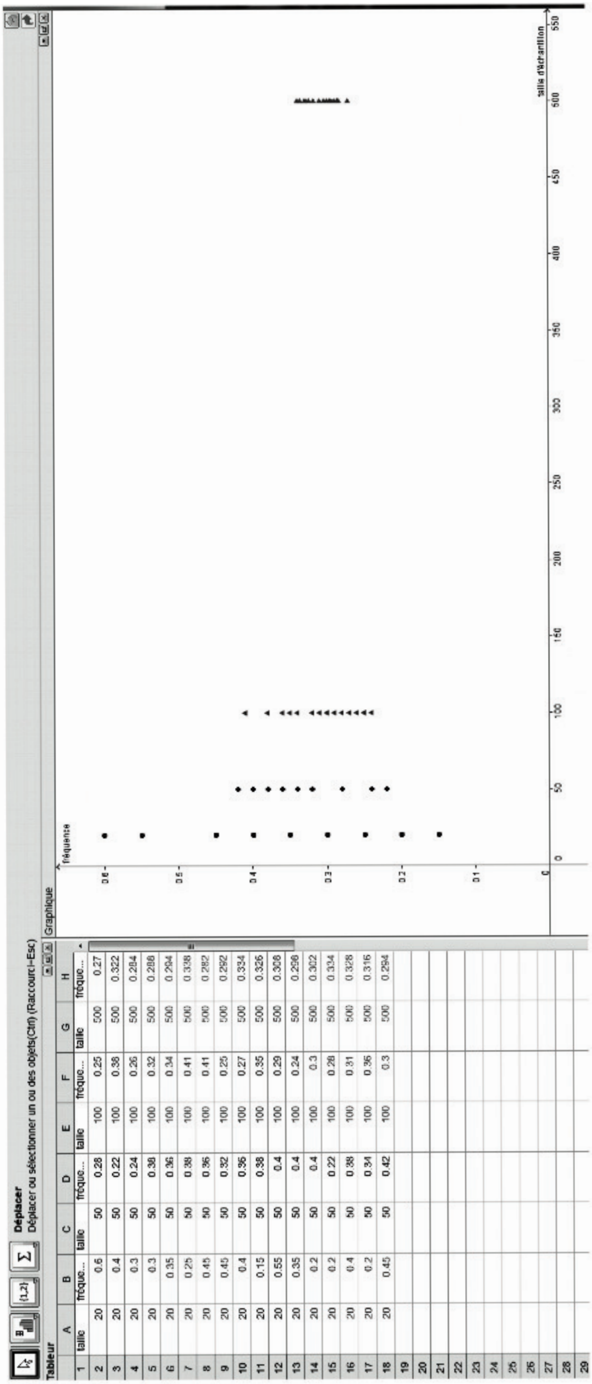
Fiche réapitulative des fréquences de boules rouges

	Echantillon A	Echantillon B	Echantillon C	Echantillon D
Nom Prénom	n = 20	n = 50	n = 100	n = 500
HANI Héloïse	0,6	0,28	0,25	0,27
Coronas Roplanté	0,4	0,22	0,38	0,322
THUAIRE . M CHARBONNE	0,3	0,24	0,26	0,284
ARRAULT I. okun/Catomet	0,3	0,38	0,32	0,288
FARTARIA - CORTIER	0,35	0,36	0,34	~~[illegible]~~ 0,294
Boulaud - GOi	0,25	0,38	0,41	0,338
Ouvry - Bourdier	0,45	0,36	0,41	0,282
Bertim - Vessaire	0,45	0,32	0,25	0,292
Bizacier Pierre	0,4	0,36	0,27	0,331
Dumoux / Mesier	0,15	0,38	0,35	0,326
Telbani / Leleau	0,55	0,4	0,29	0,308
Ruffaud / Raimbolt	0,55	0,40	0,24	0,298
Dorite / Roudaire	0,2	0,4	0,3	0,302
Cottrez / Petitjean	0,2	0,22	0,28	0,334
Valdivia / Camadon	0,4	0,38	0,31	0,328
Lacombe / Martin	0,2	0,34	0,36	0,316
Touni / Halonde	0,45	0,42	0,3	0.296

Annexe 2

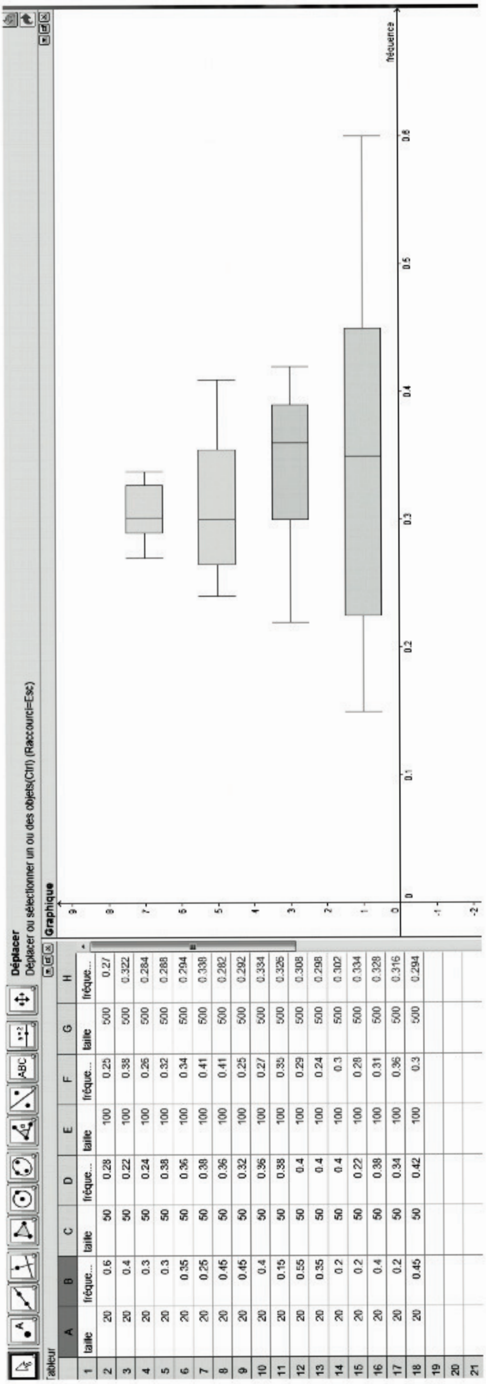

0,1
0,2
0,3
0,4
0,5
0,6
0,7
échantillon A (n = 20)
échantillon B (n = 50)
échantillon C (n = 100)
échantillon D (n = 500)
[0,05 ; 0,55]
[0,2 ; 0,4]
[0,27 ; 0,33]

Annexe 3



	A	B	C	D	E	F	G	H
1	taille	fréque...	taille	fréque...	taille	fréque...	taille	fréque...
2	20	0.6	50	0.28	100	0.25	500	0.27
3	20	0.4	50	0.22	100	0.38	500	0.322
4	20	0.3	50	0.24	100	0.26	500	0.284
5	20	0.3	50	0.38	100	0.32	500	0.288
6	20	0.35	50	0.36	100	0.34	500	0.294
7	20	0.25	50	0.38	100	0.41	500	0.338
8	20	0.45	50	0.36	100	0.41	500	0.282
9	20	0.45	50	0.32	100	0.25	500	0.292
10	20	0.4	50	0.36	100	0.27	500	0.334
11	20	0.15	50	0.38	100	0.35	500	0.326
12	20	0.55	50	0.4	100	0.29	500	0.308
13	20	0.35	50	0.4	100	0.24	500	0.298
14	20	0.2	50	0.4	100	0.3	500	0.302
15	20	0.2	50	0.22	100	0.28	500	0.334
16	20	0.4	50	0.38	100	0.31	500	0.328
17	20	0.2	50	0.34	100	0.36	500	0.316
18	20	0.45	50	0.42	100	0.3	500	0.294
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								

Annexe 4



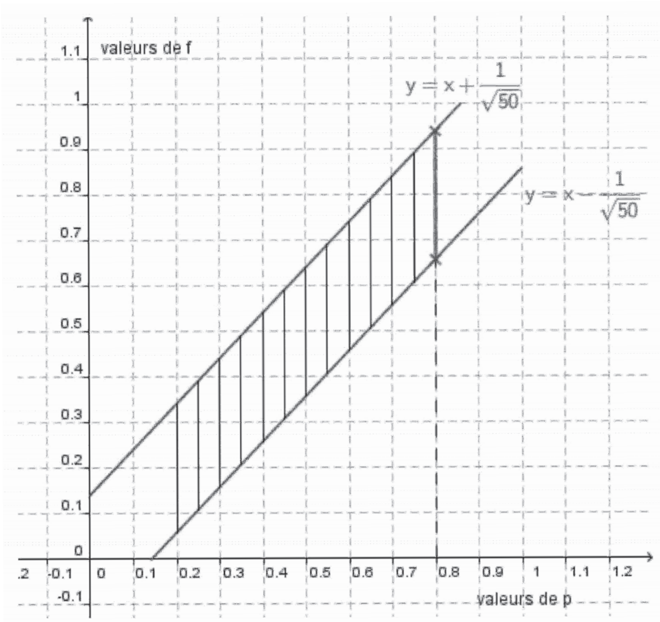
	A	B	C	D	E	F	G	H
1	taille	fréque...	taille	fréque...	taille	fréque...	taille	fréque...
2	20	0.6	50	0.28	100	0.25	500	0.27
3	20	0.4	50	0.22	100	0.38	500	0.322
4	20	0.3	50	0.24	100	0.26	500	0.284
5	20	0.3	50	0.38	100	0.32	500	0.288
6	20	0.35	50	0.36	100	0.34	500	0.294
7	20	0.25	50	0.38	100	0.41	500	0.338
8	20	0.45	50	0.36	100	0.41	500	0.282
9	20	0.45	50	0.32	100	0.25	500	0.292
10	20	0.4	50	0.36	100	0.27	500	0.334
11	20	0.15	50	0.38	100	0.35	500	0.326
12	20	0.55	50	0.4	100	0.29	500	0.308
13	20	0.35	50	0.4	100	0.24	500	0.298
14	20	0.2	50	0.4	100	0.3	500	0.302
15	20	0.2	50	0.22	100	0.28	500	0.334
16	20	0.4	50	0.38	100	0.31	500	0.328
17	20	0.2	50	0.34	100	0.36	500	0.316
18	20	0.45	50	0.42	100	0.3	500	0.294
19								
20								
21								

Annexe 5

• Algorithme pour calculer les intervalles de fluctuation au seuil de 95%, pour des valeurs de *p* allant de 0,2 à 0,8 avec un pas de 0,05 (les échantillons sont de taille *n*) :

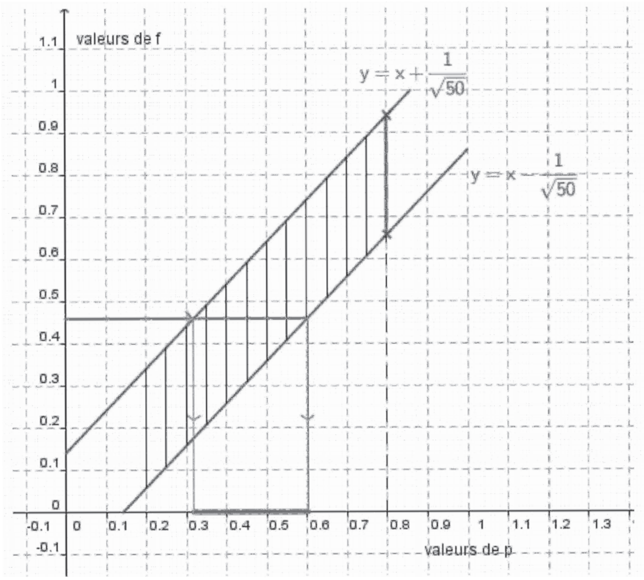
```
VARIABLES
  n EST_DU_TYPE NOMBRE
  p EST_DU_TYPE NOMBRE
  a EST_DU_TYPE NOMBRE
  b EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  AFFICHER "n = "
  LIRE n
  p PREND_LA_VALEUR 0.2
  TANT_QUE (p<=0.8) FAIRE
    DEBUT_TANT_QUE
    AFFICHER "si p = "
    AFFICHER p
    a PREND_LA_VALEUR p-1/sqrt(n)
    b PREND_LA_VALEUR p+1/sqrt(n)
    AFFICHER "IF =[ "
    AFFICHER a
    AFFICHER " ; "
    AFFICHER b
    AFFICHER " ]"
    PAUSE
    p PREND_LA_VALEUR p+0.05
    FIN_TANT_QUE
FIN_ALGORITHME
```

• Intervalles de fluctuation au seuil de 95%, pour des valeurs de *p* allant de 0,2 à 0,8 :

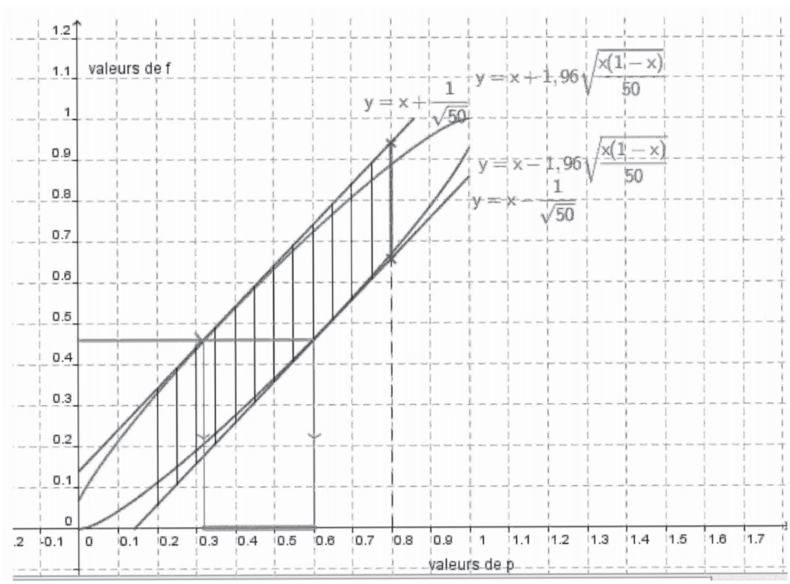


Annexe 6

• Intervalle de confiance au seuil de 95% pour une fréquence observée $f = 0{,}46$ (en utilisant l'intervalle de fluctuation vu en seconde $\left[p-\frac{1}{\sqrt{n}}\,;p+\frac{1}{\sqrt{n}}\right]$)



• Intervalle de confiance au seuil de 95% pour une fréquence observée $f = 0{,}46$ (en utilisant l'intervalle de fluctuation vu en terminale $\left[p-1{,}96\frac{\sqrt{p(p-1)}}{\sqrt{n}}\,;p+1{,}96\frac{\sqrt{p(p-1)}}{\sqrt{n}}\right]$)

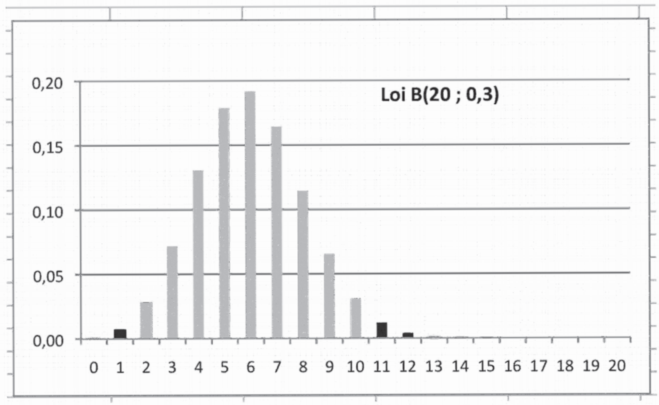


Annexe 7

Tables et représentation des lois binomiales

$\mathcal{B}$ (20 ; 0,3)

k	P(X = k)	P(X ≤ k)
0	0,0008	0,0008
1	0,0068	0,0076
2	0,0278	**0,0355**
3	0,0716	0,1071
4	0,1304	0,2375
5	0,1789	0,4164
6	0,1916	0,6080
7	0,1643	0,7723
8	0,1144	0,8867
9	0,0654	0,9520
10	0,0308	**0,9829**
11	0,0120	0,9949
12	0,0039	0,9987
13	0,0010	0,9997



$\mathcal{B}$ (50 ; 0,3)

k	P(X = k)	P(X ≤ k)
0	0,000	0,000
1	0,000	0,000
2	0,000	0,000
3	0,000	0,000
4	0,000	0,000
5	0,001	0,001
6	0,002	0,002
7	0,005	0,007
8	0,011	0,018
9	0,022	**0,040**
10	0,039	0,079
11	0,060	0,139
12	0,084	0,223
13	0,105	0,328
14	0,119	0,447
15	0,122	0,569
16	0,115	0,684
17	0,098	0,782
18	0,077	0,859
19	0,056	0,915
20	0,037	0,952
21	0,023	**0,975**
22	0,013	0,988
23	0,007	0,994
24	0,003	0,998
25	0,001	0,999