

Document 1. La fluctuation d'échantillonnage

Dans une population, lorsque l'on étudie les fréquence d'un caractère sur plusieurs échantillons de même taille d'une population, on peut observer que les résultats ne sont pas identiques selon les échantillons. Ce phénomène est appelé *fluctuation d'échantillonnage*.

Par exemple, les Crabes violonistes ont une grosse pince indifféremment à droite ou à gauche.

Si on prélève au hasard 20 *échantillons* de 10 crabes au sein de la population des Crabes violonistes, on obtiendrait des répartitions variables d'un *tirage* à l'autre.

**Document 2. Visualiser la fluctuation des fréquences liées à l'échantillonnage**

Dans une population de taille N (supposé très grand) de Crabes violonistes, on s'intéresse donc au caractère « côté de la grosse pince » dont la *modalité* est « gauche » pour 50 % des individus et « droite » pour les autres 50 %.

Pour simuler l'observation du caractère sur un échantillon de seulement 10 crabes, on peut utiliser un tableur.

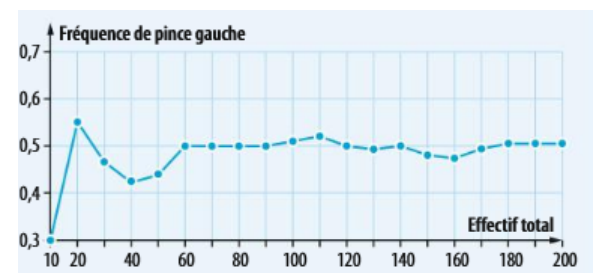
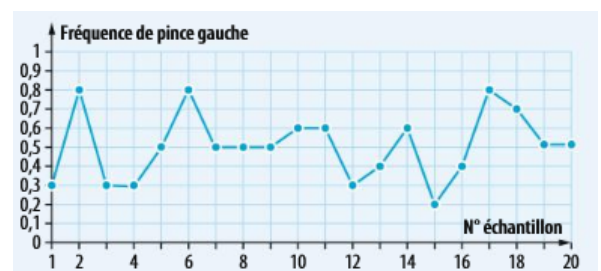
Protocole :

1. La commande `ALEA()` génère un nombre aléatoire compris entre 0 (inclus) et 1 (exclu). Pour modéliser l'équiprobabilité de la situation, on peut associer à un nombre inférieur à 0,5 le caractère « GAUCHE » et sinon le caractère sera « DROITE ». Soit `SI(ALEA()<0,5;"GAUCHE";"DROITE")`.
2. Pour le calcul de la fréquence¹ on utilise la fonction `NB.SI(B2:B11;"GAUCHE")` qui comptabilise, dans les cellules B2 à B11, combien contiennent un « GAUCHE », puis on divise par 10 pour obtenir la fréquence.

1. Si l'événement A se produit n_A fois lors d'une expérience répétée N fois, la fréquence de l'événement A est $\frac{n_A}{N}$. La fréquence est un nombre réel compris entre 0 et 1 mais on en donne souvent une valeur approchée sous forme de pourcentage.

	A	B
1	Individu	Modalité
2	1	GAUCHE
3	2	DROITE
4	3	GAUCHE
5	4	DROITE
6	5	DROITE
7	6	GAUCHE
8	7	DROITE
9	8	DROITE
10	9	DROITE
11	10	DROITE
12	Fréquence	0,3

3. En réalisant 20 fois ce tirage, on peut alors construire les représentations graphiques de la fréquence obtenue pour chaque échantillon et de la fréquence du cumul des échantillons.

**Document 3. Attribuer un niveau de confiance à une estimation de fréquence.**

La *drosophile* est une petite mouche très fréquente dans les vignobles. Elle montre des variations possibles de son phénotype : ailes développées ou non, yeux rouges ou roses, corps brun ou noir ...



Pour échantillonner les drosophiles, les scientifiques déposent de la banane écrasée dans une bouteille, qu'ils laissent ouverte une nuit : les adultes y pondent leurs œufs. Au matin, la bouteille est fermée et laissée à incuber quelques jours. Les drosophiles ainsi récoltées sont analysées.

Échantillonnage 1

Douze bouteilles sont disposées dans un vignoble bordelais. On dénombre alors 785 drosophiles dont 304 ont le corps noir, soit 38,7 % de l'échantillon.

Échantillonnage 2

Une bouteille est disposée dans 5 vignobles répartis sur le territoire français (bordeaux, pays de Loire, Bourgogne, Alsace et Languedoc). On dénombre alors en tout 168 drosophiles dont 73 sont noires, soit 43,4 % de l'échantillon.

Quelle fréquence est la plus proche de la proportion de Drosophiles noires en France ?

Aucun des deux échantillonnages n'est plus proche de la réalité : prélever au hasard sur tout le territoire est aussi approximatif que relever plus d'individus dans un seul endroit. Lorsque l'on fait ce type de sondage, on peut seulement donner un intervalle, ap-

pelé **intervalle de confiance**, dans lequel la proportion cherchée a de grandes chances de se situer. Plus l'échantillon est de grande taille et plus l'amplitude de cet intervalle sera petite ; l'estimation sera donc plus précise.

Document 4. De la fréquence d'un caractère à la fréquence d'un génotype

Pour la plupart des êtres vivants, les chromosomes sont présents par paire ; l'un est d'origine maternelle et l'autre d'origine paternelle. Les gènes² sont donc présents en deux exemplaires. Or un gène peut exister en plusieurs versions appelées allèles³.

Dans le cas de la drosophile, les individus à corps noir ont deux allèles ébène (noté e). Leur génotype est ($e // e$) : ils sont dits **homozygotes** pour e .

Les individus à corps brun possèdent au moins un allèle brun (noté B) mais peu importe le second allèle : leur génotype est soit ($B // B$), donc homozygote pour B , soit ($B // e$), donc **hétérozygote** car les deux allèles sont différents. La fréquence du génotype⁴ ($e // e$) est notée $f(e // e)$ et vaut :

$$f(e // e) = \frac{\text{nombre de drosophiles noires}}{\text{nombre total de drosophiles}}$$

Consignes.

1. **Observer** et **discuter** les fluctuations de fréquence sur les graphiques du document 2.
2. **Réaliser** une simulation en considérant cette fois 20 échantillons de 100 crabes (document 1 et protocole du document 2). **Construire** les graphiques associés les comparer à ceux du document 2.
3. **Calculer** et **comparer** les intervalles de confiance à 95% aux deux échantillons du document 3.
4. Dans l'échantillonnage 1 du document 3, on dénombre 225 homozygotes ($B // B$) dans la population récoltée. **Calculer** la fréquence du génotype ($B // e$) de cet échantillon (documents 3 et 4).

Le coin des Maths.

L'**intervalle de confiance** à 95 % est l'intervalle :

$$\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$$

où f est la fréquence observée dans un échantillon de taille n .

2. Gène : séquence ADN assurant la transmission et l'expression d'une caractéristique précise d'un organisme vivant.

3. Version possible d'un gène donné.

4. Fréquence génotypique : proportion d'un certain génotype dans une population.