En probabilités et statistiques, on approfondit le travail mené les années précédentes en l'enrichissant selon deux objectifs principaux :
- Découvrir et exploiter des exemples de lois à densité. On aborde ici le champ des problèmes à données continues. La loi uniforme fournit un cadre simple pour découvrir le concept de loi à densité et les notions afférentes. Le travail se poursuit dans le cadre des lois exponentielle et normale où le lien entre probabilité et aire est consolidé. La loi normale, fréquemment rencontrée dans les autres disciplines, doit être l'occasion d'un travail interdisciplinaire.
- Compléter la problématique de la prise de décision par celle de l'estimation par intervalle de confiance. On s'appuie sur la loi normale et, en mathématiques, on se limite au cadre d'une proportion. Toutefois, la pertinence des méthodes statistiques utilisées dans les disciplines scientifiques et technologiques, en particulier l'estimation d'une moyenne, peut s'observer par simulation.

Dans cette partie, le recours aux représentations graphiques et aux simulations est indispensable

PROGRAMMES	CAPACITES ATTENDUES	COMMENTAIRES
Prise de décision et estimation : Intervalle de fluctuation d'une fréquence	Connaître l'intervalle de fluctuation asymptotique à 95% d'une fréquence obtenue sur un échantillon de taille n : $\left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}} \,;\, p + 1{,}96\sqrt{\frac{p(1-p)}{n}} \right]$ … lorsque la proportion p dans la population est connue. Exploiter un tel intervalle pour rejeter ou non une hypothèse sur une proportion.	On fait observer que cet intervalle est proche de celui déterminé en première à l'aide de la loi binomiale, dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.
Intervalle de confiance d'une proportion	Estimer une proportion inconnue avec un niveau de confiance de 95% par l'intervalle : $\left[f - 1{,}96\sqrt{\frac{f(1-f)}{n}} \,;\, p + 1{,}96\sqrt{\frac{f(1-f)}{n}} \right]$ … calculé à partir de la fréquence f obtenue sur un échantillon de taille n. Juger de l'égalité de deux proportions à l'aide d'intervalles de confiance à 95% correspondant aux fréquences de deux échantillons de taille n.	Cette expression de l'intervalle de confiance, pour n assez grand, est admise. On constate par simulation que, pour $n \geq 30$, sur un grand nombre d'intervalles de confiance, environ 95% contiennent la proportion à estimer. La différence entre les deux fréquences observées est considérée comme significative lorsque les intervalles de confiance à 95% sont disjoints. C'est l'occasion d'étudier des méthodes statistiques pratiquées dans les disciplines scientifiques ou technologiques. En liaison avec les enseignements technologiques et scientifiques, on peut observer par simulation la pertinence d'un intervalle de confiance de la moyenne d'une population, pour un caractère suivant une loi normale. ⇔ Incertitude de mesure associée à un niveau de confiance. ⇔ Dénombrement bactérien en milieu solide.

I. ECHANTILLONNAGE

a. Population et échantillon

On considère une **population** composée d'un grand nombre d'individus, dont une certaine **proportion** p (connue) a un certain **caractère**.

On prélève un **échantillon** de n individus dans cette population. On appelle f la **fréquence** du caractère dans cet échantillon.

b. Fluctuation d'une fréquence

La **fréquence f n'est pas forcément égale à p**, tout dépendra de l'échantillon.

Mais on admettra que f suit une loi normale :
- de moyenne f
- d'écart-type $\sqrt{\frac{p(1-p)}{n}}$.

Exemple :

Dans la population, il y a 16% de gauchers (donc $p = 0{,}16$).

Je choisis un échantillon de 100 personnes au hasard (donc n = 100)

La fréquence f de gauchers dans l'échantillon suit donc une loi normale :
- de moyenne $m = 0{,}16$
- d'écart-type $\sigma = \sqrt{\frac{0{,}16 \times 0{,}84}{100}}$

On peut donc calculer la probabilité que dans cet échantillon il y ait…
… entre 10% et 30% de gauchers : $P(0{,}10 \leq f \leq 0{,}30) = 0{,}949$
… entre 15% et 20% de gauchers : $P(0{,}15 \leq f \leq 0{,}20) = 0{,}470$
… entre 15% et 17% de gauchers : $P(0{,}15 \leq f \leq 0{,}17) = 0{,}214$
Ces intervalles [0,10 ; 0,30], [0,15 ; 0,20] et [0,15 ; 0,17] sont appelés des intervalles de fluctuation de la fréquence f.

Propriété :
On appellera **intervalle de fluctuation au seuil de 95%** l'intervalle auquel f appartient avec une probabilité de 95% et qui est donné par :

$$\mathbf{I} = \left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}}\ ;\ p + 1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$$

Remarque : si on souhaitait un seuil de confiance plus élevé, on aurait un intervalle de confiance avec une amplitude telle que ça lui enlèverait tout intérêt)

c. Prise de décision

Cet intervalle de confiance sert à prendre des décisions avec un seuil de confiance de 95% (car on n'est jamais sûr de rien !) :
- on connaît la proportion p sur la population
- on connaît (ou on peut la connaître, par un simple calcul) la fréquence f dans l'échantillon.
- on sait que cette fréquence appartient à l'intervalle **I** avec une probabilité de 95%.

On fait **l'hypothèse** suivante : *l'échantillon provient bien de la population* (« est représentatif »)
Si f appartient bien à I, alors on accepte l'hypothèse. Sinon on la rejette.

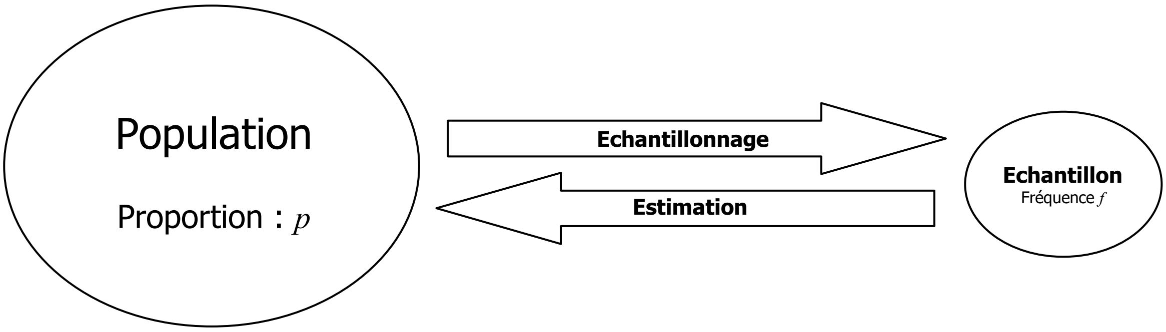
Exemple :
Lors d'une élection, 25% de la population a voté pour Mr Blanc.
On dispose d'un échantillon (qu'on suppose représentatif de la population) de 150 personnes dans lequel 45 personnes ont voté pour Mr Blanc.

On a donc : $p = 0{,}25$ $\quad n = 150$ $\quad f = \dfrac{45}{150} = 0{,}30$ $\quad \sqrt{\dfrac{p(1-p)}{n}} = 0{,}035$
On calcule l'intervalle de confiance $[0{,}25 - 1{,}96 \times 0{,}035\ ;\ 0{,}25 + 1{,}96 \times 0{,}035] = [0{,}181\ ;\ 0{,}319]$
On a bien $f \in [0{,}181\ ;\ 0{,}319]$ donc **on accepte l'hypothèse**.

II. ESTIMATION ET INTERVALLE DE CONFIANCE

a. Estimation et échantillonnage

Jusqu'ici, on a fait étudié les fluctuations de la fréquence f sur un échantillon, connaissant la proportion p sur une population : c'est de **l'échantillonnage**.
Inversement, lorsqu'on utilise les informations de l'échantillon pour en déduire des informations sur la population, c'est de **l'estimation** (concrètement, c'est le cas de sondages)



Propriété :

Si dans un échantillon de taille n, un caractère a une fréquence d'apparition f, alors la proportion de ce caractère p dans l'ensemble de la population appartient à un certain intervalle avec une probabilité de 95%, qu'on appelle intervalle de confiance à 95% et qui est donné par :

$$\mathbf{I} = \left[f - 1{,}96\sqrt{\frac{f(1-f)}{n}}\,;\, f + 1{,}96\sqrt{\frac{f(1-f)}{n}}\right]$$

Exemple :

On a interrogé un échantillon de 997 personnes, dont 274 ont déclaré avoir l'intention de voter pour Mr Blanc.

On a donc : $n = 997$ $\quad f = \dfrac{274}{997} = 0{,}275$ $\quad \sqrt{\dfrac{f(1-f)}{n}} = 0{,}014$

On peut donc dire que sur l'ensemble de la population, il y a 95 de chances que la proportion de gens votant pour Mr Blanc appartienne à l'intervalle : $[0{,}275 - 1{,}96 \times 0{,}014\ ;\ 0{,}275 - 1{,}96 \times 0{,}014]$ = **[0,248 ; 0,302]**

b. Comparaison de deux proportions

On souhaite comparer deux échantillons.

On pose l'hypothèse : *ces deux échantillons proviennent de la même population.*

On procède de la façon suivante :

- On calcule les fréquences f_1 et f_2 sur chacun des deux échantillons.
- On détermine les deux intervalles de confiance à 95% correspondants.

Si les deux intervalles sont disjoints (aucune intersection) on rejette l'hypothèse. Sinon on l'accepte.