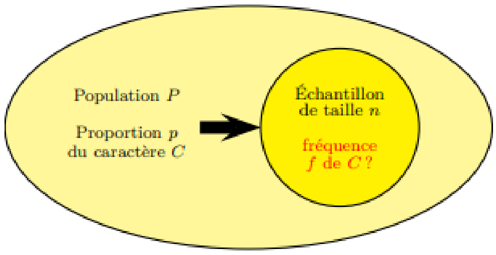
Chapitre PS4 : Echantillonnage & Estimation

1 – Intervalle de fluctuation de la fréquence (Echantillonage)



● Soit une population, dont la **proportion** p d'un caractère est connue.

Ex : Dans la population française, il y a 13 % de gauchers.

● On prélève un échantillon de **taille** n dans cette population et on s'intéresse à la **fréquence** f du caractère dans cet échantillon.

Ex : Votre classe est un échantillon de taille $n =$ …… et la fréquence de gauchers est $f =$ ……………………………

● Si l'on évalue la fréquence f de ce caractère dans plusieurs échantillons de même taille n, on obtiendra des résultats légèrement différents. Ce phénomène s'appelle **fluctuation d'échantillonnage.**

Ex : Dans une autre classe, la fréquence de gauchers pourra être différente.

● Plus la taille n des échantillons est grande, moins les variations de f seront importantes : La fréquence f observée dans les échantillons se rapproche de la proportion p.

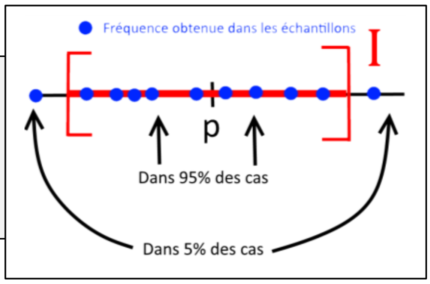
Ex : Résultats obtenus sur 10 classes de 40 élèves puis sur 10 amphithéâtres de 250 élèves :

Classe	1	2	3	4	5	6	7	8	9	10
Gauchers	10,00%	32,50%	10,00%	7,50%	10,00%	15,00%	10,00%	2,50%	7,50%	12,50%

Amphi	1	2	3	4	5	6	7	8	9	10
Gauchers	13,60%	11,60%	10,80%	18,40%	12,00%	11,20%	13,40%	10,40%	17,50%	14,40%

● Afin de rendre compte des fluctuations naturelles de la fréquence observée dans les différents échantillons on utilise un intervalle :

Définition 1 : On dit que I est **intervalle de fluctuation au seuil de 95%**, s'il contient la fréquence f observée dans les échantillons de taille n, dans au moins 95% de cas.



Propriété 1 : Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ alors $I = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ est un intervalle de fluctuation au seuil de 95%.

Exemple 1 : Calculer l'intervalle de fluctuation de la fréquence de gauchers dans des échantillons de taille n

● Comme $p = 0.13$ et que $np > 5$ on doit avoir $n \times 0.13 > 5$ c'est-à-dire $n > \frac{5}{0.13} \approx 38{,}49$ donc $n \geq 39$

● $n = 40$: $I = \left[0.13 - \frac{1}{\sqrt{40}} ; 0.13 + \frac{1}{\sqrt{40}}\right] = [0.13 - 0.158\,; 0.13 + 0.158] = [0; 0.288]$

● $n = 250$: $I = \left[0.13 - \frac{1}{\sqrt{250}} ; 0.13 + \frac{1}{\sqrt{250}}\right] = [0.13 - 0.063\,; 0.13 + 0.063] = [0.067; 0.193]$

Remarque : Plus n est grand, plus l'intervalle sera précis.

2 – Prise de décision

Soit une population dans laquelle on suppose l'hypothèse : H : « La proportion d'un caractère est p ».

On prélève un échantillon de taille n et on note f la fréquence observée.

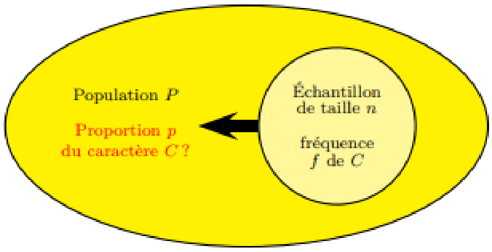
Règle de décision : Si I est un intervalle de fluctuation au seuil de 95 %. Alors :

- Si $f \in I$, alors on accepte l'hypothèse avec un seuil de confiance de 95 %
- Si $f \notin I$, alors on rejette l'hypothèse avec un risque de 5 %

Exemple 3 : On suppose l'hypothèse que le pourcentage de gauchers dans le monde est aussi de 13%. On prélève un échantillon de 250 personnes dans le monde et on dénombre 28 gauchers. Doit-on accepter ou rejeter l'hypothèse ?

$f = \frac{28}{250} = 0.112 \in [0.067; 0.193]$. On accepte donc l'hypothèse avec un seuil de confiance de 95%.

3 – Intervalle de confiance de la proportion (Estimation)



- Soit une population, dont la **proportion** p d'un caractère est inconnue.

Ex : Avant le vote (2016), quelle est la proportion de britannique qui vont voter pour le BREXIT ?

- On prélève un échantillon de **taille** n dans cette population et on s'intéresse à la **fréquence** f du caractère dans cet échantillon.

Ex : On effectue un sondage sur 1000 britanniques choisis au hasard dans la population parmi lesquels 512 se sont prononcés en faveur du BREXIT. Dans ce cas on a $n = 1000$ et $f = 0.512$.

- Problématique : Peut –on en conclure que le BREXIT aura lieu ?

Pour répondre à cette question, on utilise un intervalle dit de confiance.

Définition 1 : On dit que I est **intervalle de confiance au seuil de 95%**, si la proportion p appartient à cet intervalle avec une probabilité de 95%.

Propriété 3 : Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ alors l'intervalle $I = \left[f - \frac{1}{\sqrt{n}} \,;\, f + \frac{1}{\sqrt{n}}\right]$ est un intervalle de confiance de la proportion p.

Remarque : Il s'agit de la même formule que pour l'intervalle de fluctuation mais avec f au lieu de p.

Exemple 3 : Calculons l'intervalle de confiance de la proportion de britanniques qui voteront pour le BREXIT :

$$I = \left[0.512 - \frac{1}{\sqrt{1000}} \,;\, 0.512 + \frac{1}{\sqrt{1000}}\right] = [0.48; 0.54]$$

Avec un seuil de confiance à 95%, on peut dire la proportion de britanniques qui se prononcera pour le BREXIT sera comprise entre 48% et 54% (si les intentions de votes restent les mêmes).

Loi binomiale, Loi Normale – Fiche d'exercices

Intervalle de fluctuation

2 Dans une population, la proportion d'un caractère est $p = 0,6$.
Pour un échantillon de 200 individus, déterminer l'intervalle de fluctuation de la fréquence, à 95 %.
Approcher les bornes à 0,01 près.

6 CALC Dans chaque cas, déterminer l'intervalle de fluctuation, à 95 %, de la fréquence pour la proportion p donnée, dans un échantillon de taille n donné.

1. $p = 0,45$ et $n = 64$
2. $p = 0,7$ et $n = 400$
3. $p = 0,5$ et $n = 1\,600$

8 CALC En décembre 2012, sur Télérama, une enquête indique que 40 % des personnes de 15 ans et plus ont regardé des programmes de télévision sur un écran d'ordinateur ou une tablette.

Calculer l'intervalle de fluctuation, à 95 %, de la fréquence pour un échantillon de taille 800.

15 CALC Un sondage indique que 43 % de la population ont déjà utilisé un écran autre que le téléviseur pour regarder la télévision.
Le directeur d'une chaîne de télévision a supervisé une enquête auprès de 400 téléspectateurs d'une émission d'information et parmi eux 165 ont utilisé un écran autre que le téléviseur pour visionner cette émission.
Peut-il rejeter la proportion de 43 % proposée par l'institut de sondage ?

Intervalle de confiance

27 CALC Un groupe d'assurances établit des prévisions pour l'année à venir.
Le responsable considère un échantillon de 400 véhicules assurés depuis au moins un an dans ce groupe.

1. Il constate que 340 véhicules de cet échantillon n'ont pas eu de sinistre.
a) Déterminer l'intervalle $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ où f est la fréquence des véhicules n'ayant pas eu de sinistre et n la taille de l'échantillon considéré.
b) Interpréter concrètement le résultat.

2. Lequel des intervalles semble le plus correct pour estimer le pourcentage de véhicules qui auront un sinistre, au niveau de confiance à 95 %.
a) entre 15 % et 25 % b) entre 10 % et 20 %
c) entre 80 % et 90 %

30 CALC Dans une ville de 32 400 ménages, la proportion p de ménages d'une ville qui possèdent deux voitures est inconnue.
Pour un échantillon de 500 ménages, 30 % d'entre eux possèdent deux voitures.
Donner une estimation de p par l'intervalle de confiance, au niveau de confiance à 95 %.
Indiquer la fourchette de cette estimation en pourcentage, puis en nombre de ménages approché à 10 ménages près.

29 CALC Le directeur d'une grande entreprise propose un bilan de compétences à ses salariés.
Il prépare son projet par un sondage.
Sur 900 salariés interrogés, 180 sont intéressés.

1. Déterminer l'intervalle $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$, où f est la fréquence des salariés intéressés pour faire ce bilan et n le nombre de salariés interrogés.
2. La direction générale de l'entreprise souhaite que ce bilan de compétences concerne 30 % des salariés.
D'après l'intervalle de confiance obtenu en **1.**, le souhait de la direction générale a-t-il beaucoup de chance de se réaliser ?

Exercices type bac

Ex 1 QCM (Tiré de divers sujets de bac)

1. La part de consommateurs bio réguliers, c'est-à-dire ceux qui disent consommer bio au moins une fois par mois s'élève à 43 % en France en 2015.
On effectue un sondage dans une société de 400 personnes.
La fréquence de consommateurs bio réguliers dans cet échantillon est notée f.
a) $f = 0{,}43$ b) Au seuil de 95 %, $0{,}38 \leq f \leq 0{,}48$
c) Au seuil de 95 %, $0{,}4275 \leq f \leq 0{,}4325$ d) Au seuil de 95 %, $0{,}23 \leq f \leq 0{,}63$

2. Une chaîne d'hôtels internationale annonce un taux de remplissage de ses chambres de 74,8% pour l'année 2015 dans le monde. Cette année là, cette chaîne possédait 850 chambres en France.
Un intervalle de fluctuation au seuil de 95% du taux de remplissage des chambres en France pour 2015 est :
a. $[74{,}80;74{,}84]$ *b.* $[0{,}747;0{,}749]$ *c.* $[0{,}713;0{,}783]$ *d.* $[0{,}715;0{,}780]$

3. Dans une population, on estime qu'il naît 51% de garçons et 49% de filles.
L'intervalle de fluctuation à au moins 95% de la fréquence des filles dans un échantillon de 100 naissances choisies au hasard sera :
a. $[0{,}48 ; 0{,}50]$ **b.** $[0{,}39 ; 0{,}59]$ **c.** $[0{,}41 ; 0{,}61]$ **d.** $[0{,}47 ; 0{,}51]$

Ex 2 QCM 2 (Tiré de divers sujets de bac)

1. On procède à un contrôle technique de 100 scooters constituant un échantillon représentatif des scooters circulant dans une ville.
27 de ces scooters sont déclarés en mauvais état.
À partir de ce résultat, on souhaite estimer la proportion de scooters en mauvais état circulant dans la ville.
Un intervalle de confiance, au niveau de confiance de 95 %, pour la proportion de scooters en mauvais état dans la ville est :
a) $[0{,}26 ; 0{,}28]$ b) $[0{,}2 ; 0{,}3]$ c) $[0{,}17 ; 0{,}37]$ d) $[0{,}27 ; 0{,}95]$

2. Un maire souhaite estimer la proportion d'habitants de sa commune satisfaits des décisions qu'il a prises depuis son élection. Un récent sondage effectué sur 800 habitants montre que 560 personnes sont satisfaites.
Un intervalle de confiance au niveau de confiance 95 % pour la proportion d'opinions favorables est :
a. $[0{,}66 ; 0{,}74]$ **b.** $[0{,}69 ; 0{,}71]$ **c.** $[0{,}60 ; 0{,}80]$ **d.** $[0{,}71 ; 0{,}79]$

Ex 3 Football (Tiré du bac Pondichery 2014)

En France, en 2011, 22 % des sportifs licenciés avaient une licence de football.
Déterminer un intervalle de fluctuation à au moins 95 % de la fréquence des licenciés de football dans un échantillon de 400 sportifs licenciés choisis au hasard parmi les sportifs licenciés en 2011.

Ex 4 Laboratoire pharmaceutique (Tiré du bac Antilles-Guyanne 2017)

Un laboratoire pharmaceutique fabrique un traitement préventif pour éviter la survenue de cette maladie. Il avertit que 30 % des patients traités auront des effets secondaires.

Plusieurs études sont réalisées par différents médecins et des patients volontaires pour vérifier les estimations du laboratoire. Les médecins sont invités à rentrer leurs données dans un logiciel qui utilise l'algorithme ci-contre :

1. Un médecin a traité 150 patients ; parmi ceux-ci, 40 ont eu des effets secondaires. Quel sera le résultat affiché par ce logiciel ?

2. Pour un autre, sur 200 patients, 75 ont eu des effets secondaires. Qu'affichera alors le logiciel ?

3. Que représente dans cet algorithme l'intervalle $[a ; b]$?

Variables :
    n, s sont des entiers
    a, b sont des nombres réels
Entrée :
    Afficher « Entrer le nombre de patients traités »
    Saisir n
    Afficher « Entrer le nombre de patients ayant eu des effets secondaires »
    Saisir s
Traitement :
    a prend la valeur $0{,}3 - \frac{1}{\sqrt{n}}$
    b prend la valeur $0{,}3 + \frac{1}{\sqrt{n}}$
    Si $a \leq \frac{s}{n} \leq b$
        Alors afficher : « résultats conformes »
        Sinon afficher : « résultats non conformes »
    Fin Si

Ex 5 Laboratoire pharmaceutique (Tiré du bac Métropole Septembre 2017)

Le comité d'une entreprise a récemment fait construire une salle de sport pour améliorer le bien-être de ses salariés.
Le président du comité déclare que 80 % des salariés sont satisfaits de la qualité des nouvelles installations sportives. Alix mène une enquête auprès de 300 de ses collègues choisis au hasard. Parmi eux, 228 se déclarent satisfaits des installations sportives.

1. Déterminer un intervalle de fluctuation, au seuil de 95 %, de la proportion de personnes satisfaites dans cet échantillon. Arrondir les bornes au centième.

2. Les résultats de l'enquête menée par Alix peuvent-ils remettre en question les propos du président ?

Ex 6 **Parc aquatique (Tiré du bac Métropole 2015)**

Pour le repas du midi, les visiteurs restant toute la journée dans le parc peuvent :

- soit déjeuner dans l'un des restaurants du parc ;
- soit consommer, sur une aire de pique-nique, un repas qu'ils ont apporté.

La direction souhaite estimer la proportion p de visiteurs déjeunant dans l'un des restaurants du parc.

Un sondage est effectué à la sortie du parc : 247 visiteurs parmi 625 ont déjeuné dans l'un des restaurants du parc.

Déterminer un intervalle de confiance au niveau de confiance de 95 % de la proportion p de visiteurs déjeunant dans l'un des restaurants du parc.

Ex 7 **Plaques métalliques (Tiré du bac Pondichery 2018)**

Pour la fabrication de machines agricoles, une usine reçoit en grande quantité des plaques métalliques carrées. Elles ne peuvent être utilisées dans le processus de fabrication que si la longueur de leurs côtés et leur épaisseur respectent certains critères.

1. Un premier test permet de vérifier la longueur des côtés de chaque plaque. Une plaque réussit ce test si la longueur de ses côtés est comprise entre 81,6 centimètres et 82,4 centimètres. On note X la variable aléatoire qui, à chaque plaque prélevée au hasard, associe la longueur de son côté, en centimètres. On admet que la variable aléatoire X suit la loi normale d'espérance 82 et d'écart-type 0,2.

   Déterminer la probabilité, arrondie au millième, qu'une plaque réussisse ce premier test.

2. Les plaques ayant réussi le premier test subissent un second test permettant de vérifier leur épaisseur. Une plaque sera utilisable par l'usine si son épaisseur est inférieure à 3 millimètres. Le fournisseur affirme que 90 % des plaques qui subiront ce second test ont une épaisseur inférieure à 3 millimètres.
   On effectue le second test sur un lot de 2 500 plaques.

   a. Déterminer l'intervalle de fluctuation, à au moins 95 %, de la fréquence des plaques dont l'épaisseur est inférieure à 3 millimètres, dans ce lot.

   b. Parmi les 2 500 plaques, 2 274 ont réussi le second test. Au regard de ces résultats, doit-on accepter l'affirmation du fournisseur ?

Ex 8 **Blanchisserie (Tiré du bac Centres Etrangers 2018)**

Une entreprise de blanchisserie fait réaliser une étude par une société de conseil spécialisée dans l'accompagnement vers la transition énergétique.

La société de conseil affirme au gérant que 90 % des clients sont sensibles aux questions environnementales.

Avant de remplacer son parc de machines, le gérant réalise un sondage auprès de 350 clients.

Ce sondage révèle alors que, parmi eux, 290 y sont sensibles.

Ce résultat permet-il de remettre en cause l'affirmation de la société de conseil ?

Argumenter la réponse à l'aide d'un intervalle de fluctuation.

Ex 9 **Hotline (Tiré du bac Polynésie 2015)**

Une société de hotline fait une enquête sur le niveau de satisfaction des personnes qui ont recours à leurs services par téléphone. Elle dispose de deux centres d'appel : un situé à Marseille, un autre situé à Lille.

L'enquête consiste à demander à chaque personne ayant téléphoné si elle est satisfaite ou non du service que la hotline lui a proposé.

La société estime que 58 % des appels reçus l'ont été par le centre de Marseille.

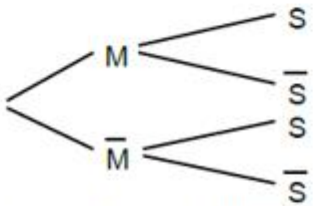
De plus, parmi les appels reçus par le centre de Marseille, on constate un taux de 34 % de personnes satisfaites ; alors que pour le centre de Lille, on constate un taux de 44 % de personnes satisfaites.

On choisit au hasard une personne ayant téléphoné. On considère les évènements suivants :

M : « la personne a téléphoné au centre de Marseille ».

S : « la personne est satisfaite du service proposé ».

1. **a.** Recopier et compléter l'arbre de probabilité suivant :



b. Déterminer la probabilité que la personne ait téléphoné au centre de Marseille et soit satisfaite.

c. Montrer que la probabilité que la personne ayant téléphoné soit satisfaite est $p = 0,382$.

2. Sachant que la personne ayant téléphoné a été satisfaite, quelle est la probabilité que cette personne ait téléphoné au centre de Lille ?
Arrondir le résultat à 0,001 près.

3. On considère un échantillon de 500 personnes choisies au hasard ayant téléphoné à l'un des centres d'appel.
Déterminer un intervalle de fluctuation au seuil de 95 % du taux de personnes satisfaites pour cet échantillon. *Arrondir les bornes à 0,001 près.*

Ex 10 Elections (Tiré du bac Polynésie 2017)

Des sondages quotidiens ont été effectués avant le second tour d'une élection opposant deux candidats A et B. Les intentions de votes, en pourcentage, pour le candidat A sont données dans le tableau suivant :

Dates :	24/04	25/04	26/04	27/04	30/04	01/05	02/05	03/05	04/05
Rang du jour x_i	1	2	3	4	7	8	9	10	11
Pourcentage y_i	55	55	54,5	55	54	53,5	53	53	52

Par exemple, le 24 avril les intentions de votes pour le candidat A étaient de 55% et pour le candidat B de 45%.

Le scrutin aura lieu le 6 mai. Comme il est interdit de publier des résultats de sondages les deux derniers jours avant le scrutin, on ne dispose pas des sondages pour le 5 et le 6 mai.

Le nuage de points de coordonnées $(x_i\,;y_i)$ pour i variant de 1 à 11, est donné en annexe 1 à rendre avec la copie.

1. À l'aide de la calculatrice, déterminer, par la méthode des moindres carrés, une équation de la droite d'ajustement de y en x (arrondir les coefficients au millième).
2. On décide d'ajuster le nuage avec la droite D d'équation $y = -0{,}28x + 55{,}6$.
   - **a)** Tracer la droite D sur le graphique figurant sur annexe.
   - **b)** Déterminer la valeur prévue par ce modèle le 6 mai, jour de l'élection.
   - **c)** Si l'élection n'avait pas eu lieu le 6 mai, d'après ce modèle, à partir de quelle date le candidat B serait-il passé en tête des sondages ?
3. Des sondages ont été faits le jour de l'élection mais n'ont pas été communiqués. Un de ces sondages donnait le candidat A à 52 %. L'institut disait avoir effectué ce sondage sur un échantillon représentatif de 1225 personnes.
   - **a)** Au vu de ce dernier sondage, établir l'intervalle de confiance au niveau de 95%, pour le résultat du candidat A à l'élection.
   - **b)** Au vu de cet intervalle, la victoire de ce candidat-semblait elle assurée ? *Justifier la réponse.*