Échantillonnage & Estimation

Vue d'ensemble : Résumé de toute une scolarité d'échantillonnage et d'estimation en un coup d'œil

- Soit p la proportion réelle dans la <u>population totale</u> = la <u>probabilité</u> qu'un individu choisi au hasard dans la population ait la caractéristique à laquelle on s'intéresse. C'est la fréquence <u>théorique</u>.
- Soit f la fréquence <u>observée</u> de cette caractéristique dans <u>l'échantillon</u>.

	Échantillonnage ***p connue ou bien on veut tester une hypothèse sur la valeur de p*** Utilisation = Prise de décision à partir d'un échantillon : On regarde si la valeur observée pour f est « raisonnable » *(Pour voir s'il y a discrimination contre les femmes dans une entreprise, pour voir si un jury a vraiment été choisi au hasard...etc)*	**Estimation** ***p inconnue** (et on n'a aucune hypothèse sur la valeur de p)* Utilisation = Estimation : On estime p au moyen de f. *(pour estimer la proportion p de gens qui vont voter pour un candidat à partir des résultats d'un sondage...etc)*
En seconde	Si $n \geqslant 25$ et $0{,}2 \leqslant p \leqslant 0{,}8$ alors pour au moins 95% des échantillons, la fréquence **observée *f*** vérifie $f \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$ *Intervalle de **fluctuation** au seuil de 95%*	Si $n \geqslant 25$ et $0{,}2 \leqslant f \leqslant 0{,}8$ alors pour au moins 95% des échantillons, la fréquence **théorique** vérifie $p \in \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$ *Intervalle de **confiance*** au seuil de 95%

Le tirage au hasard dans la population d'un individu qui peut présenter le caractère C avec la probabilité p est une épreuve de Bernoulli de paramètre p où le succès est l'issue « avoir C ». Le prélèvement au hasard d'un échantillon de taille n dans cette population s'assimile à un schéma de Bernoulli. La variable aléatoire qui compte le nombre de succès, c'est à dire le nombre d'individus présentant le caractère C, suit la loi binomiale $\mathscr{B}(n, p)$. Voilà pourquoi les méthodes de Première et Terminale se focalisent sur les loi binomiales.

En 1S (si on a le temps de finir le programme)	On calcule les probabilités d'avoir k succès sur n tirages pour la loi binomiale et on en déduit l'intervalle de fluctuation au seuil de 95%.	
En TS *(seulement le Best of)*	Si $n \geqslant 30$, $n\,p \geqslant 5$ et $n(1-p) \geqslant 5$, alors pour environ 95% des échantillons, on peut considérer que la fréquence **observée *f*** vérifie $f \in \left[p - 1{,}96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1{,}96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$ ↑ ***Intervalle de fluctuation asymptotique au seuil de 95%*** (Le 1,96 vient de la loi normale)	*Même résultat qu'en seconde* : Pour au moins 95% des échantillons, la fréquence **théorique *p*** vérifie $p \in \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$ ↑ ***Intervalle de confiance au seuil de 95%*** mais sous les conditions $n \geqslant 30$, $n\,f \geqslant 5$ et $n(1-f) \geqslant 5$
Un jour peut-être		*Résultat symétrique de celui de TS sur l'intervalle de fluctuation :* Pour environ 95% des échantillons, la fréquence **théorique *p*** vérifie $p \in \left[f - 1{,}96 \frac{\sqrt{f(1-f)}}{\sqrt{n}} ; f + 1{,}96 \frac{\sqrt{f(1-f)}}{\sqrt{n}} \right]$ *Intervalle de confiance au seuil de 95%* ↥ sous les conditions $n \geqslant 30$, $n\,f \geqslant 5$ et $n(1-f) \geqslant 5$

<u>Remarque</u> : Nous ne mentionnerons plus jamais le dernier résultat mais vu la symétrie des résultats d'échantillonnage et d'estimation vus en seconde, on ne peut pas s'empêcher de se demander si le résultat d'échantillonnage de TS n'aurait pas son symétrique en estimation : La réponse est donc « *oui, mais le résultat en question n'est pas su programme de TS* ». Savoir que ce résultat existe satisfait cependant notre sens de l'esthétique.