# Estimations statistiques

**Définition** Une expérience aléatoire est une expérience pour laquelle il est impossible de prévoir le résultat, qui est soumis .......................... Par exemple, le lancer d'un dé ou le tirage d'une boule dans une urne.

Si l'on souhaite répéter une expérience un très grand nombre de fois, comme lancer 5 000 fois un dé, on peut faire soi-même l'expérience avec un dé ou utiliser un simulateur comme .......................................

**Définition** Un .................... de taille $n$ est la série statistique formée des $n$ résultats obtenus lorsqu'on répète $n$ fois une même expérience dans les mêmes conditions.
La distribution des fréquences est la liste des ......................... d'un échantillon.

**Propriété** Les distributions des fréquences .................... d'un échantillon à l'autre pour une même expérience : c'est ce qu'on appelle la **fluctuation d'échantillonnage**. Lorsque la taille $n$ de l'échantillon augmente l'ampleur des fluctuations diminue et les fréquences ont tendances à se stabiliser.

## 1. Intervalle de fluctuation

**Définition** En statistique, on peut s'intéresser à un caractère qui est soit présent soit non présent (être une femme dans une population humaine, obtenir une somme égale à 7 dans un lancer de deux dés). Si l'on appelle $p$ la probabilité que le caractère soit présent, ainsi la probabilité qu'il ne le soit pas est ....................
On peut dire que la population est compatible avec un modèle de Bernoulli de probabilité $p$.

**Propriété** On admet que si on a un échantillon de taille $n$ d'une population, dont un caractère a la probabilité $p$ d'être présent et tel que $n \geq 25$ et $0.2 \leq p \leq 0.8$, alors dans ce cas pour 95% des échantillons la fréquence d'apparition du caractère appartient à l'intervalle

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

**Définition** Cet intervalle est appelé **intervalle de fluctuation au seuil de** 95%.

## 2. Prendre une décision à partir d'un échantillon

Pour décider si une fréquence $f$ observée sur un échantillon de taille $n$ est compatible ou non avec un modèle de Bernoulli de probabilité $p$ on teste l'appartenance de $f$ à l'intervalle de fluctuation au seuil de 95% donné ci-dessus. Par suite,

- Si f n'est pas dans l'intervalle de fluctuation, alors on peut .................... l'hypothèse que l'échantillon est compatible avec le modèle.
- Si f est dans l'intervalle de fluctuation, alors on .................... l'hypothèse que l'échantillon est compatible avec le modèle.

**Remarque** Quelle que soit la décision prise il y a toujours le risque que ce ne soit pas la bonne décision.

**Exercice 1** Dans une usine automobile, on contrôle les défauts de peinture de type « grains ponctuels sur le capot ». Lorsque le processus est sous contrôle, on a 20% de ce type de défauts. Lors du contrôle aléatoire de 50 véhicules, on observe 26% de défauts (13 sur 50). Faut-il s'inquiéter ?

## 3. Estimer une proportion inconnue p à partir d'un échantillon

D'après la propriété il y a 95% de chance pour que $f \in \left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$, ce qui peut s'écrire qu'au seuil de confiance de 95% on a

$$p \in \left[ \qquad ; \qquad \right]$$

Ainsi, connaissant un échantillon aléatoire de taille $n$ dont la fréquence $f$ est connue on peut estimer la proportion $p$ du caractère dans la population par un intervalle appelé **intervalle de confiance** avec une probabilité d'au moins 0.95.

**Exercice 2** Le taux d'audience d'une chaîne de télévision, un soir à 20H, a été mesuré à partir de 1 000 appareils : on a trouvé 31%. Donner une estimation en pourcentages de l'audience de cette chaîne ce soir-là pour le pays entier sous forme d'intervalle.