

# Chapitre VI

## Échantillonnages et simulations

### 1. INTERVALLE DE FLUCTUATION AU SEUIL DE 95%

On fait un relevé statistique de taille  $n$  sur une population pour laquelle la probabilité d'avoir le caractère étudié est  $p$ .

Lorsque  $n \geq 25$  et  $0,2 \leq p \leq 0,8$ , il y a au moins 95% d'échantillons de taille  $n$  dont la fréquence  $f$  pour le caractère étudié est dans l'intervalle

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

Cet intervalle est appelé l'intervalle de fluctuation au seuil de 95%.

#### Application : Acceptation ou rejet d'un échantillon

Pour décider si un échantillon de taille  $n$  est représentatif d'un modèle ayant une probabilité de  $p$ , on détermine l'intervalle de fluctuation du modèle puis on teste l'appartenance de la fréquence  $f$  observée à cet intervalle.

- Si  $f$  n'est pas dans l'intervalle, alors on peut rejeter l'échantillon au prétexte qu'il n'est pas compatible avec le modèle.
- Si  $f$  est dans l'intervalle alors on estime que l'échantillon est compatible avec le modèle.

*Remarque :* Quelque soit la décision prise, il y a un risque que ce ne soit pas la bonne décision dans 5% des cas.

### 2. INTERVALLE DE CONFIANCE AU SEUIL DE 95%

On fait un relevé statistique sur un échantillon (d'une population). La fréquence observée sur cet échantillon pour un caractère donné est  $f$ .

Il y a 95% de chances que la probabilité du caractère sur la population soit dans l'intervalle

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

#### Application : Estimation d'une probabilité

Pour estimer la probabilité  $p$  d'un caractère dans une population à partir de la fréquence  $f$  fournie par un échantillon de taille  $n$ , on utilise l'intervalle de confiance

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

Cet intervalle contient  $p$  au niveau de confiance de 95%.

### 3. CONCLUSION

Il y a deux démarches très différentes.

- Soit on connaît la probabilité  $p$  du caractère étudié et dans ce cas on utilise l'intervalle de fluctuation au seuil de 95% pour savoir si l'échantillon testé est représentatif de la population totale.
- Soit on ne connaît pas  $p$  et dans ce cas, plutôt que de baser sur une estimation ponctuelle, on calcule l'intervalle de confiance au seuil de 95%. Ainsi on sait qu'il y a 95% de chance que  $p$  soit dans l'intervalle  $\left[ f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}} \right]$

### 4. GLOSSARY

English	French	
Survey	Sondage	A method for collecting quantitative information about items in a population.
Sample	Échantillon	A subset of a population selected for measurement, observation or questioning, to provide statistical information about the population.
Sampling	Échantillonnage	The process or technique of obtaining a representative sample.
Sampling fluctuation	Fluctuation d'échantillonnage	Refers to the extent to which a statistic takes on different values with different samples.
Estimate (verb)	Estimer	To calculate roughly, often from imperfect data.
Estimate	Estimation	A rough calculation or guess.
Estimation	Estimation	The process of making an estimate.
Margin of error	Marge d'erreur	An expression of the lack of precision in the results obtained from a sample.
Point estimate	Estimation ponctuelle	A single value computed from sample data, used as a "best guess" for an unknown population parameter.
Confidence interval	Intervalle de confiance	A particular kind of interval estimate of a population parameter.
Simulate (verb)	Simuler	To model, replicate, duplicate the behavior, appearance or properties of a system or environment
Simulation	Simulation	Something which simulates a system or environment in order to predict actual behaviour.