CHAPITRE 13

Échantillonnage

Plan de ce chapitre

I. Introduction

Nous allons considérer l'expérience consistant à lancer un dé cubique supposé non truqué.
On pose alors :

A : « obtenir le 1 ».

Nous allons effectuer une simulation de 100 lancers de dé à l'aide d'un tableur.
Ouvrons un tableur et entrons dans la cellule `A1` la formule suivante :

	A	B	C
1	`=ENT(1+6*ALEA())`		
2			

- La fonction `ALEA()` permet de calculer un nombre aléatoire compris entre 0 (compris) et 1 (non compris) ;
- `6*ALEA()` affichera donc un nombre dans l'intervalle `[0;6[` ;
- `1+6*ALEA()` affichera donc un nombre dans l'intervalle `[1;7[` ;
- La fonction `ENT` calcule la partie entière d'un nombre; donc `ENT(1+6*ALEA())` affichera un nombre entier parmi : 1, 2, 3, 4, 5 et 6.

Maintenant, copions cette formule de `A2` à `A100` afin d'obtenir, au final, 100 nombres entiers compris entre 1 et 6.
Dans la cellule `A101`, entrons la formule :

`=NB.SI(A1:A100;1)`

Elle affichera le nombre de « 1 » qui apparaissent dans la colonne `A`.
Ensuite, entrons dans la cellule `A102` la formule :

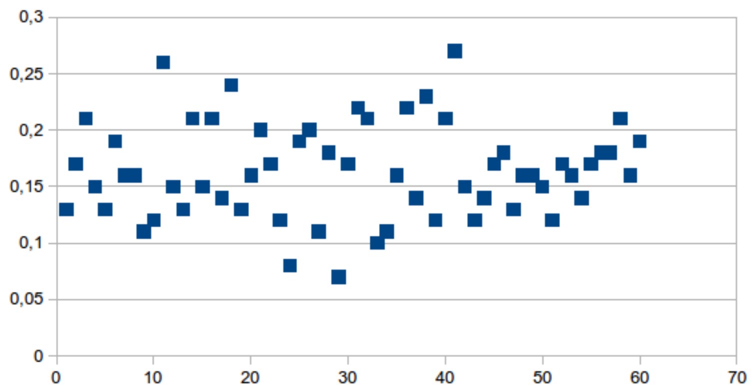
`=A101/100`

pour afficher la fréquence d'apparition du nombre « 1 » dans la colonne `A`.

Copions alors la colonne `A` dans les colonnes de `B` à `BG` afin d'obtenir au final 60 colonnes de nombres.
Ceci simule 60 échantillons de 100 lancers d'un dé cubique (60 échantillons de taille $n = 100$).

On construit alors un nuage de points représentant les différentes fréquences obtenues.

En appuyant sur la touche `[F9]`, on rafraîchit les calculs et on obtient un autre graphique.
Voici un des graphiques obtenus :



La fréquence d'apparition théorique du nombre « 1 » est la probabilité d'obtenir ce nombre :

$$p = \frac{1}{6} \approx 0,17.$$

On s'aperçoit alors sur le graphique que les fréquences semblent se regrouper autour de cette valeur. On peut donc imaginer que les fréquences d'apparition du nombre « 1 » vont se trouver dans un intervalle centré en p : un intervalle de la forme $\mathrm{I} = \left[p - \delta ; p + \delta\right]$.
Nous allons prendre $\delta = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{100}} = 0,1$. Alors, $\mathrm{I} \approx [0,066 ; 0,267]$.
Pour connaître le pourcentage de fréquences comprises dans cet intervalle, insérons dans la cellule `A103` la formule :

```
=SI(A102<=0,266;SI(A102>0,067;1;0);0)
```

qui affiche « 1 » si tel est le cas, et « 0 » sinon. Ensuite, copions-la sur toute la ligne 103 jusqu'à `BH103`.
Dans la cellule `A104`, insérons la formule :

```
=SOMME(A103:BH103)/60
```

puis formatons cette cellule de sorte à ce qu'elle affiche un pourcentage (en cliquant sur l'icône « % ») : ceci affichera le pourcentage de fréquences d'apparition du « 1 » qui sont dans I.

En appuyant sur la touche `[F9]` plusieurs fois, je vois les pourcentages suivants :

$$100\,\% \quad ; \quad 96,67\,\% \quad ; \quad 98,33\,\% \quad ; \quad 100\,\% \quad ; \quad 100\,\% \quad ; \quad 98,33\,\%.$$

En utilisant donc cette valeur de δ, nous sommes *a priori* assurés qu'au moins 95 % des fréquences se trouvent dans I.

II. Intervalle de fluctuation

II. 1. Échantillon

Déf.1 Soit une population dans laquelle on observe un groupe de n individus.
Ce groupe est appelé un **échantillon de taille n** de la population.

Exemple 1

Dans l'expérience précédente, qui consiste à lancer 100 fois un dé cubique, l'échantillon était de taille 100 car on a effectué 100 lancers de dés (que l'on a ensuite répété 60 fois).

II. 2. Intervalle de fluctuation au seuil de 95%

Définition 2

On considère une expérience aléatoire dans laquelle on considère un événement A de probabilité p. On effectue n fois cette expérience ; on note alors f la fréquence d'apparition de l'événement A sur cet échantillon de taille n.
Un **intervalle de fluctuation au seuil de** 95% de f, noté I_f, est un intervalle centré en p qui contient f dans un échantillon de taille n avec une probabilité égale à 0,95 :

$$\mathrm{I}_f = \left[p - \delta\,; p + \delta\right] \quad (\delta > 0) \qquad \text{et} \qquad \mathrm{P}(f \in \mathrm{I}_f) = 0{,}95.$$

À retenir

Un intervalle de fluctuation n'est pas unique car il y a plusieurs valeurs possible de δ.

Propriété 1

Considérons un échantillon de taille n et posons p la probabilité d'obtenir un résultat précis. Notons f la fréquence d'apparition de ce résultat dans l'échantillon.
Pour $0{,}2 \leqslant p \leqslant 0{,}8$ et $n \geqslant 25$, un intervalle de fluctuation au seuil de 95% est :

$$\mathrm{I}_f = \left[p - \frac{1}{\sqrt{n}}\,; p + \frac{1}{\sqrt{n}}\right].$$

II. 3. Prise de décision

On utilise les intervalles de fluctuations pour déterminer si l'on peut considérer qu'un événement est « normal » ou pas.

Exemple 2

On lance 100 fois un dé tétraédrique (4 faces) et on observe que la fréquence d'apparition de la face « 3 » est $f = 0{,}32$. On peut donc se demander si le dé n'est pas truqué car la fréquence observé est plus grande que la probabilité de l'événement, à savoir $p = \frac{1}{4} = 0{,}25$.
Un intervalle de fluctuation au seuil de 95% est :

$$\mathrm{I}_f = \left[0{,}25 - \frac{1}{\sqrt{100}}\,; 0{,}25 + \frac{1}{\sqrt{100}}\right] = [0{,}15\,; 0{,}35].$$

On constate alors que $f \in \mathrm{I}_f$, ce qui nous permet de dire que le dé n'est probablement pas truqué.

En revanche, si $f = 0{,}29$ sur 10 000 lancers, l'intervalle de fluctuation devient :

$$\mathrm{I}'_f = \left[0{,}25 - \frac{1}{\sqrt{10\,000}}\,; 0{,}25 + \frac{1}{\sqrt{10\,000}}\right] = [0{,}24\,; 0{,}26].$$

Dans ce cas, $f \notin \mathrm{I}'_f$, ce qui nous permet de dire que le dé est probablement truqué.