Adéquation à une loi équirépartie

Une *loi équirépartie* est une loi uniforme d'une variable aléatoire X qui peut prendre n valeurs de telle sorte que la probabilité soit la même pour chacune de ces n valeurs.

<u>Problème</u> : *Un joueur veut vérifier si le dé qu'il possède est « normal », c'est-à-dire bien équilibré.*
On sait que, dans ce cas-là, la loi de probabilité associée est la loi uniforme :

$$P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = \frac{1}{6}$$

Pour cela, le joueur lance 200 fois le dé et note les résultats obtenus :

x_i	1	2	3	4	5	6
n_i	31	38	40	32	28	31
f_i	0,155	0,190	0,200	0,160	0,140	0,155

Pour savoir si la distribution de fréquences obtenue est « proche » de la loi uniforme, on calcule la quantité suivante, qui prend en compte l'écart existant entre chaque fréquence trouvée et la probabilité théorique attendue :

$$d^2 = \left(0{,}155 - \frac{1}{6}\right)^2 + \left(0{,}190 - \frac{1}{6}\right)^2 + \left(0{,}200 - \frac{1}{6}\right)^2 + \left(0{,}160 - \frac{1}{6}\right)^2 + \left(0{,}140 - \frac{1}{6}\right)^2 + \left(0{,}155 - \frac{1}{6}\right)^2.$$

(On note d^2 car son calcul est celui du carré d'une distance)
Mais rien ne permet de dire pour l'instant si cette quantité trouvée est « petite » ou « grande ». En effet, elle est soumise à la fluctuation d'échantillonnage, puisque sa valeur varie d'une série de lancers à l'autre. On va donc étudier cette fluctuation d'échantillonnage pour convenir d'un seuil entre « petite » et « grande » valeur de d^2 lorsqu'on lance 200 fois un dé. Pour cela, on génère des séries de 200 chiffres au hasard pris dans $\{1;2;3;4;5;6\}$. Les résultats trouvés pour le nombre d^2 à partir de 1 000 simulations sont résumés par le tableau suivant :

Minimum	$\mathbf{D_1}$	$\mathbf{Q_1}$	**Médiane**	$\mathbf{Q_3}$	$\mathbf{D_9}$	**Maximum**
0,000363	0,00138	0,00233	0,00363	0,00555	0,00789	0,01658

Le **neuvième décile** de la série des valeurs simulées de d^2 est 0,00789.
Cela signifie que 90 % des valeurs de d^2 obtenues au cours de ces 1 000 simulations sont dans l'intervalle [0;0,00789]. Comme la valeur observée de d^2 est inférieure à cette valeur seuil de 0,00789, on peut convenir que le dé est équilibré avec un risque de 10 %.
En effet, en utilisant cette méthode sur les données simulées, on se serait trompé dans 10 % des cas. On dit que l'on a un **seuil de confiance** de 90 %.

<u>Propriété</u> : Soit une épreuve conduisant aux issues $a_1, a_2, .., a_q$. Expérimentalement, si on répète n fois cette épreuve ($n \geqslant 100$) , on obtient les fréquences $f_1, f_2 \dots. f_q$ pour chacune de ces issues.
Vérifier l'adéquation de ces données à la loi équirépartie sur $\{a_1, a_2, \dots a_q\}$ c'est calculer

$$d^2 = \sum_{i=1}^{q} \left(f_i - \frac{1}{q}\right)^2$$

Sachant que D_9 est le neuvième décile :

- Si $d^2 \leqslant D_9$ alors on dira que les données sont compatibles avec le modèle de la loi uniforme au seuil de risque de 10 %.
- Si $d^2 > D_9$ alors on dira que les données ne sont pas compatibles avec le modèle de la loi uniforme au seuil de risque de 10 %.