Adéquation de données expérimentales à une loi équirépartie

Texte proposé, dans le cadre du groupe académique « Lycée », par Jean-Raymond Delahaye - Lycée Alain Borne - Montélimar - jeanray.delahaye@wanadoo.fr

Le problème posé

On a réalise l'expérience suivante : lancer 200 fois un dé.
Les fréquences f_1, f_2, f_3, f_4, f_5, f_6, obtenues sont les suivantes :

1	2	3	4	5	6
0,195	0,11	0,19	0,195	0,16	0,15

Le modèle théorique associé au dé équilibré donne la loi de probabilité suivante :

1	2	3	4	5	6
0,1667	0,1667	0,1667	0,1667	0,1667	0,1667

On observe évidemment des fluctuations entre les fréquences et la loi de probabilité.

Comment définir un critère qui permette de décider si le dé est pipé ou non ?

Pour mesurer l'écart entre la distribution des fréquences et la loi de probabilité, on utilise :

$$d^2 = (f_1 - 1/6)^2 + (f_2 - 1/6)^2 + (f_3 - 1/6)^2 + (f_4 - 1/6)^2 + (f_5 - 1/6)^2 + (f_6 - 1/6)^2 .$$

Dans notre exemple, $d^2 = 0,00568$.

Ce réel d^2 fluctue d'un échantillon à l'autre, c'est normal ! La question qui se pose est la suivante :

À partir de quelle valeur dira-t-on que d^2 est trop grand pour que l'on puisse imputer l’écart à la fluctuation d'échantillonnage ?
Si on observe que ce seuil est dépassé, on décidera que le dé est pipé.
Sinon, on décidera que le dé est équilibré.

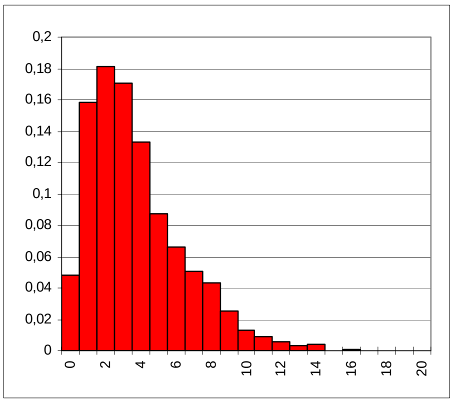
La détermination du seuil

On fait l'hypothèse que le dé est équilibré.
L’écart d^2 suit une loi de probabilité qui peut être étudiée de façon théorique. En terminale *S*, on ne dispose pas des outils nécessaires. On va donc utiliser une méthode expérimentale.

On simule 1000 séries de 200 lancers de ce dé.

On obtient donc une série de 1000 valeurs de d^2.

A droite figure l'histogramme des fréquences de d^2 (par commodité, on a déterminé les fréquences de 1000 d^2 après regroupement en classes d'amplitude 1).



Dans cet échantillon de 1000 valeurs de 1000 d^2, le neuvième décile est 7,48.

Cela signifie que, dans cet échantillon,

10% des valeurs de 1000 d^2 sont supérieures à 7,48 ;
90% des valeurs de 1000 d^2 sont inférieures à 7,48.

On décide de prendre comme seuil de décision cette valeur 7,48.

La conclusion

Dans l'expérience réalisée, on a trouvé 1000 d^2 = 5,68.
Comme 5,68 < 7,48, on déclare que le dé est équilibré (on accepte l'hypothèse "le dé est équilibré").

En procédant ainsi, on risque de rejeter à tort l'hypothèse dans 10% des cas.
Avec les mêmes valeurs expérimentales, on pouvait tester et accepter d'autres hypothèses.
On ne démontre donc pas que le dé est équilibré.

La formule donnant l'écart entre « distribution des fréquences » et « loi de probabilité » dans les manuels

Ω contient r issues.

La loi de probabilité est définie par les p_i .

On simule n fois l'expérience.

La distribution des fréquences empiriques est définie par les f_i .

Selon les manuels, on trouve les formules suivantes :

- $d^2 = \sum_{1}^{r} (f_i - p_i)^2$
- $D^2 = n \sum_{1}^{r} \frac{(f_i - p_i)^2}{p_i}$

Si la loi de probabilité est équirépartie, pour tout i, $p_i = p$.

Alors $D^2 = n \sum_{1}^{r} \frac{(f_i - p_i)^2}{p_i} = \frac{n}{p} \sum_{1}^{r} (f_i - p_i)^2 = \frac{n}{p} d^2$.

Le neuvième décile est donc le même dans les deux cas.

Si la loi est équirépartie, les deux distances d et D sont donc équivalentes.

Quelques informations complémentaires à propos de $D^2 = n \sum_{1}^{r} \frac{(f_i - p_i)^2}{p_i}$

La variable D^2 suit une loi du Khi-deux.

Cette loi ne dépend pas de n (si n >30), elle ne dépend que de r.

E(D^2) = r–1 (nombre de degrés de liberté)

Les tables donnent les valeurs de p($D^2 < a$) selon les valeurs de a et r.