ADEQUATION A UNE LOI EQUIREPARTIE

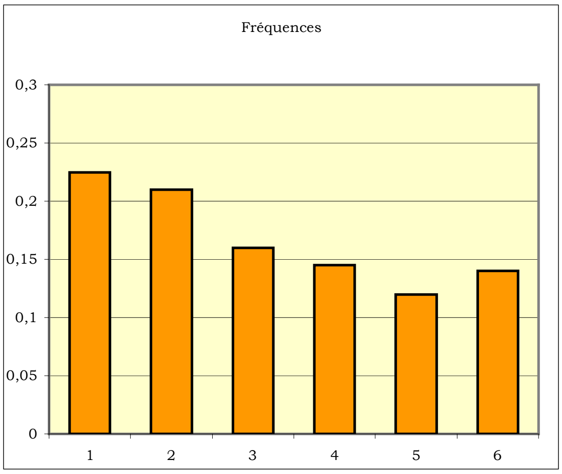
1 Introduction :

Comment, à partir d'un échantillon de données expérimentales liées à un phénomène aléatoire, décider, par une argumentation de nature statistique, qu'un modèle mathématique est en **adéquation** avec la réalité?
Avant de jouer avec un dé, on aimerait savoir s'il est bien équilibré, c'est à dire si chaque face a autant de "chances" de sortir; le modèle mathématique sous jacent est l'équiprobabilité.

Pour vérifier ce modèle, on peut jouer 200 fois avec le dé, construire une statistique de cet échantillon, en faire le traitement (représentation graphique, paramètres de position et de dispersion), puis comparer avec les données théoriques, en évaluant les écarts.

Exemple : On trouve sur une série de 200 lancers :

Faces	Effectifs	Fréquences f_i	Valeur Théorique
1	45	0,225	$\frac{1}{6}$
2	42	0,21	$\frac{1}{6}$
3	32	0,16	$\frac{1}{6}$
4	29	0,145	$\frac{1}{6}$
5	24	0,12	$\frac{1}{6}$
6	28	0,14	$\frac{1}{6}$

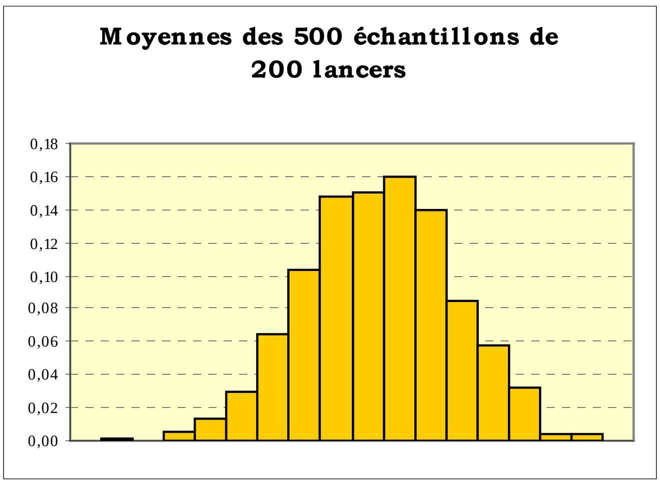


2 Étude de la moyenne :

La moyenne de cette série est 3,145 : elle est différente de celle attendue avec un dé parfait : 3,5.

Pour savoir si cette différence est significative, on effectue une **simulation** d'une loi équirépartie sur {1 ; 2 ; 3 ; 4 ; 5 ; 6} à l'aide d'un tableur : la fonction Ent(6 × alea() + 1) permet ainsi de simuler un dé parfait.

On simule, par exemple, 500 échantillons de taille 200 de ce "dé parfait". Une étude statistique des moyennes de ces 500 échantillons permet de mieux cerner les **fluctuations d'échantillonnage** des échantillons de taille 200.



On constate statistiquement que la répartition des moyennes des 500 échantillons donne un histogramme en forme de cloche : 5 % des moyennes sont inférieures à 3,3 et 5 % sont supérieures à 3,7; donc 90 % des moyennes sont entre 3,3 et 3.7.

On adopte la règle de décision suivante : si la moyenne observée dans l'échantillon fabriqué à partir du dé n'est pas dans cet intervalle, on décide de dire qu'il y a "peu de chance" que le dé soit parfait.

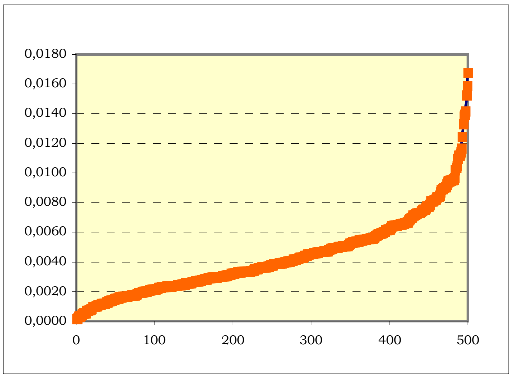
Pour l'instant le dé a "moins de 10 % de chance" d'être parfait puisque la moyenne trouvée 3,145 n'est pas dans cet intervalle.

Sur la série de 200 lancers du dé, on étudie la "distance observée" donnée par

$$d_{obs}^2 = \sum_{i=1}^{i=6}\left(f_i - \frac{1}{6}\right)^2$$

La distance observée 0,008 683 n'est pas nulle; pour savoir si cette distance est la conséquence d'un dé mal équilibré, ou si elle est due à une fluctuation d'échantillonnage, on utilise de nouveau la simulation : l'étude statistique des distances calculées à partir de 500 échantillons simulés de taille 200 permet de mieux cerner les fluctuations d'échantillonnage des distances.

Faces	Fréquences f_i	Valeur Théorique	Distances $(f_i - 1/6)^2$
1	0,225	1	0,003 403
2	0,21	1	0,001 878
3	0,16	1	0,000 044
4	0,145	1	0,000 469
5	0,12	1	0,002 178
6	0,14	1	0,000 711
Total			0,008 683



On constate statistiquement que 90 % des distances sont inférieures à 0,008 (le 9ème décile vaut 0,008).

On adopte la règle de décision suivante : si la distance observée dans l'échantillon fabriqué à partir du dé est plus grande que 0,008, on décide de dire qu'il y a peu de chance que le dé soit parfait.

4 Conclusion : Avec ce critère, le dé a moins de 10 % de chance d'être parfait et sera déclaré non conforme au modèle d'équiprobabilité.

Remarques : Le seuil choisi ici en tant que critère de décision est de 10%. En procédant ainsi, on risque de rejeter à tort l'hypothèse dans 10% des cas.

Avec les mêmes valeurs expérimentales et un seuil différent, on peut tester et accepter l'hypothèse. On ne démontre donc pas que le dé est équilibré.

Critères de décision :

- Si $d_{obs}^2 < D_9$, on accepte le modèle d'équirépartition, c'est-à-dire que l'on considère que les résultats observés sont compatibles avec l'hypothèse que le dé est bien équilibré.
- Si $d_{obs}^2 > D_9$, on rejette l'hypothèse que le dé est équilibré **au risque d'erreur de 10%.**

5 Adéquation à une loi équirépartie :

On considère une épreuve ayant k issues équiprobables (6 dans le cas du lancer de dé) répétée n fois, avec n assez grand (supérieur à 30). On effectue une simulation dans les conditions vues précédemment, on étudie la somme des écarts au carré entre les fréquences et la valeur théorique $\frac{1}{k}$, et on choisit un seuil de référence selon la précision souhaitée :

Si l'on choisit D9 (cas fréquent) : cela signifie que peu de valeurs (10%) lui sont supérieures ; en conséquence :

Lorsque $d_{obs}^2 > D_9$ **on rejette l'hypothèse d'équirépartition au risque d'erreur de 10%**

Lorsque $d_{obs}^2 < D_9$ on peut accepter l'hypothèse d'équirépartition mais sans certitude, ni même évaluation du risque d'erreur.

Remarques :

(1) Par ces simulations, on ne prouve pas que le dé est équilibré. On fait une "étude comparative" qui donne une idée plus ou moins valide de la réponse.

(2) Dans le cas où on rejette l'hypothèse d'équirépartition, on évalue le risque de se tromper : si l'on choisit D9, le dé est considéré comme non équilibré avec une marge d'erreur de 10%. Une marge d'erreur de 1% correspond au 99ème centile : cela signifie qu'on considère que seulement 1% des résultats de la simulation sont marginaux. Ici, si on choisit le seuil de 1%, comme le 99ème centile vaut 0,014, le dé étudié est considéré comme équilibré : on ne prend donc pas le risque de dire que le dé est déséquilibré, mais on n'est pas sûr pour autant qu'il soit équilibré !