

TEST D'ADÉQUATION

Séminaire IREM du 20 juin 2003

Groupe de travail : statistiques et probabilités
IREM de Besançon.

Sommaire

1	Introduction	11
2	Étude de la moyenne	12
3	Étude de la distance séparant les fréquences observées de 1/6	12
4	Conclusion	13

1 Introduction

Le propos est d'illustrer sur un exemple une démarche développée par le groupe « statistiques et probabilité », le 5 mars 2003 devant les professeurs de mathématiques de l'Académie. Comme le dit si justement Yves Ducl : « Comment, à partir de l'observation d'un échantillon, décider par une argumentation de nature statistique qu'un modèle mathématique est en adéquation avec la réalité ? »

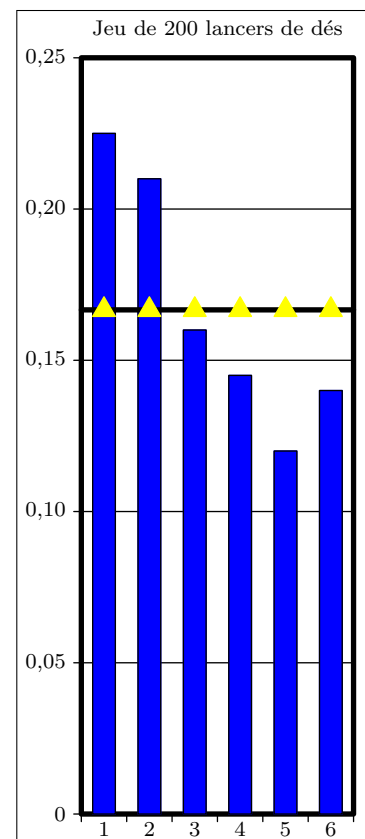
Prenons un dé.

Avant de jouer avec ce dé on aimerait savoir s'il est bien équilibré. Dans cette question, il y a un modèle mathématique qui se cache : chaque face a autant de chance de sortir, on dit que les événements sont « équiprobables ».

Pour vérifier ce modèle on peut jouer 200 fois avec le dé, construire une statistique de cet échantillon, et en faire le traitement habituel : représentation graphique, paramètres de position et de dispersion.

On trouve :

Faces	Effectifs	Fréquences	Modèle	Distance
1	45	0,225	1/6	0,003 403
2	42	0,21	1/6	0,001 878
3	32	0,16	1/6	0,000 044
4	29	0,145	1/6	0,000 469
5	24	0,12	1/6	0,002 178
6	28	0,14	1/6	0,000 711
	200			0,008 683



La moyenne des points dans l'échantillon est de 3,145.

La distance entre les fréquences observées et celles attendues est

$$d^2 = \sum_{i=1}^6 \left(f_i - \frac{1}{6} \right)^2 \simeq 0,008683$$

2 Étude de la moyenne

La moyenne observée 3,145 est différente de celle espérée du dé parfait : 3,5.

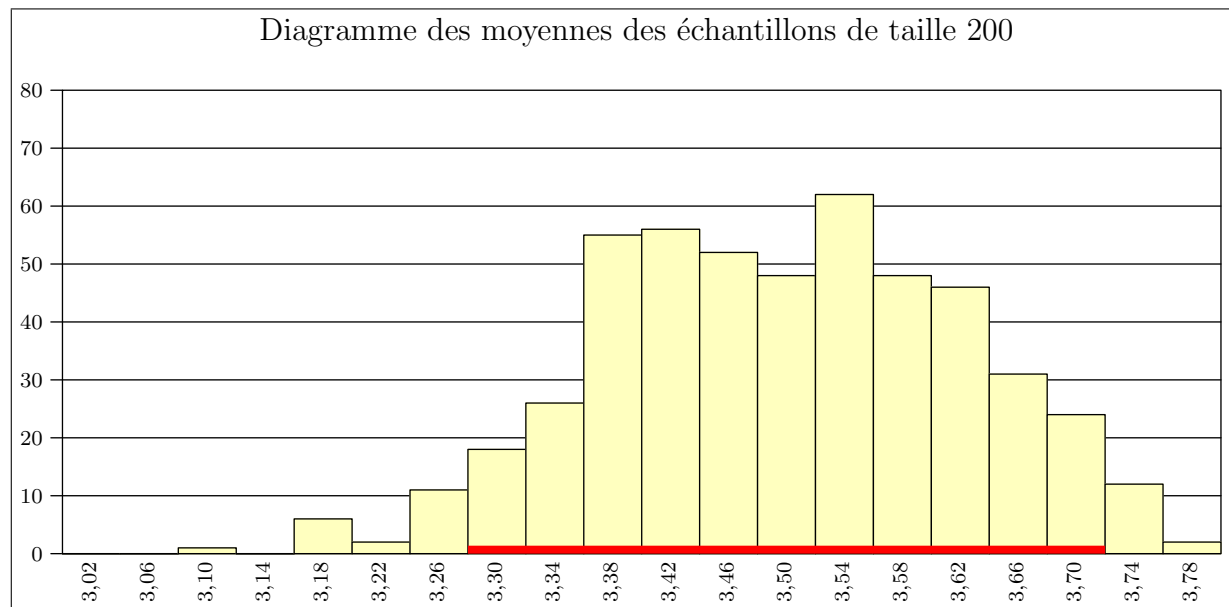
Pour savoir si cette différence est significative ou si elle est la conséquence d'une fluctuation d'échantillonnage, utilisons la méthode suivante :

Avec un tableur et la fonction `Int(6xAléat+1)` on simule un dé parfait.

On construit 500 échantillons de taille 200 de ce dé parfait.

Une étude statistique (graphique, paramètres) des moyennes de ces échantillons nous permet de mieux cerner les fluctuations d'échantillonnage des moyennes des échantillons de taille 200.

On constate statistiquement que la répartition des moyennes a un histogramme en forme de cloche, que 5 % des moyennes sont inférieures à 3,3, 5 % sont supérieures à 3,7 donc que 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 de mon dé n'est pas dans cet intervalle je décide de dire qu'il y a peu de chance que mon dé soit parfait.

Pour l'instant mon dé a moins de 10 % de chance d'être parfait.

3 Étude de la distance séparant les fréquences observées de $1/6$

La distance observée 0,008 683 n'est pas nulle.

Pour savoir si cette distance est la conséquence d'un dé mal fait ou si elle est due à une fluctuation d'échantillonnage, utilisons la même méthode :

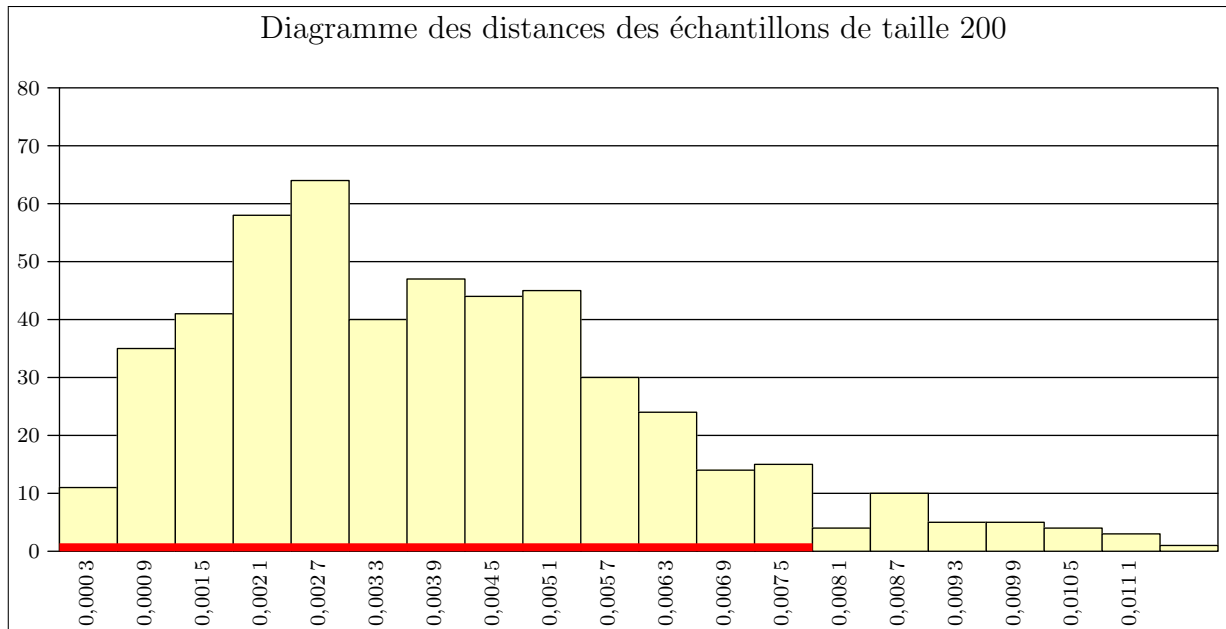
Une étude statistique (graphique, paramètres, déciles) des distances calculées à partir des échantillons simulés de taille 200 nous permet de mieux cerner les fluctuations d'échantillonnage des distances.

Test d'adéquation

On constate statistiquement que l'histogramme des distances a une forme dissymétrique très particulière.

On constate que 90 % des distances sont inférieures à 0,008.

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



Avec ce critère, mon dé a moins de 10 % de chance d'être parfait.

Je refuse donc de croire que mon dé est conforme au modèle.

4 Conclusion

Cette démarche est l'occasion d'étudier, à partir d'un caractère discret des variables continues, de les observer statistiquement pour reconnaître des modèles mathématiques que les élèves pourront étudier plus tard.

Michel Vendrely