

COMMENT DETECTER UNE PIECE TRUQUEE

d'après Simulation et Statistique en Seconde, IREM Paris Nord, 2000
et TD 5 de Transmath, TS, 2002

I) Fluctuations de la fréquence d'apparition de pile, selon le nombre de lancers

Si l'on possède une pièce de monnaie bien équilibrée, la "loi des grands nombres" veut que, plus on fait de lancers, plus la fréquence d'apparition de "pile" s'approche de $1/2$:

On effectue une série de 100 lancers d'une pièce et on calcule la fréquence \hat{p} d'apparition de pile :

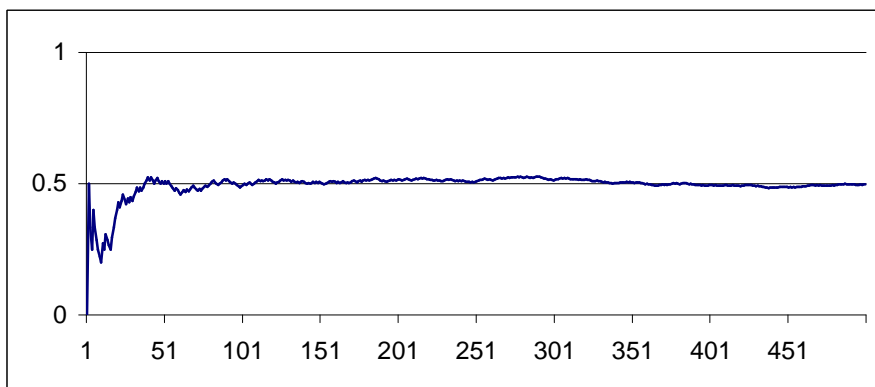
élèves	1	2	3	4	5	6	7	8	9	10
fréquences de pile	0,51	0,51	0,43	0,52	0,57	0,51	0,57	0,47	0,52	0,47

La moyenne des fréquences est la fréquence \hat{p} d'apparition de pile, observée sur les 100 lancers est 0,508

De façon à augmenter sans peine le nombre de lancers, on va recourir à la simulation.

Le programme suivant permet de visualiser, sur 500 lancers, l'évolution de la **proportion** (fréquence) de "piles".

CASIO Graph 25 → 100	T.I. 80 - 82 - 83	T.I. 89 - 92
ViewWindow 0,500 ,100,0.4,0.6,0.1↵ Graph Y= 0.5↵ 0 → N ↵ For 1 → I To 500 ↵ Int(Ran#+0.5) + N →N↵ Plot I , N ÷ I ↵ Next ↵	:FnOff :PlotsOff :0 → Xmin :500 → Xmax :100 → Xscl :0.4 → Ymin :0.6 → Ymax :0.1 → Yscl :DrawF 0.5 ::0 → N :For (I , 1 , 500) :int(rand+0.5) + N → N :Pt-On (I , N / I) :End	:FnOff :ClrDraw :PlotsOff :0 → xmin :500 → xmax :100 → xscl :0.4 → ymin :0.6 → ymax :0.1 → yscl :DrawFunc 0.5 ::0 → n :For i , 1 , 500 :int(rand()+0.5) + n → n :PtOn i , n/i :EndFor



Pour cette partie, on admet que, pour 500 lancers d'une pièce non truquée, on a 95 % de chances d'observer une fréquence \hat{p} d'apparition de pile comprise entre 0,456 et 0,544.

Rappel du résultat de seconde :

Si la probabilité p se trouve entre 0,30 et 0,70, dans plus de 90% des cas, le fréquence observée se situe dans l'intervalle $[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}]$; résultat que l'on applique ici pour $p = \frac{1}{2}$ et $n = 500$.

Modifions le programme précédent de façon à expérimenter ce test en modifiant aléatoirement la valeur de p . La calculatrice choisit aléatoirement la valeur de p (fréquence théorique d'apparition de pile) par l'instruction : $0.5+(-1)^{\text{int}(\text{rand}+0.5)} \times 0.06 \times \text{int}(\text{rand}+0.5) \rightarrow p$

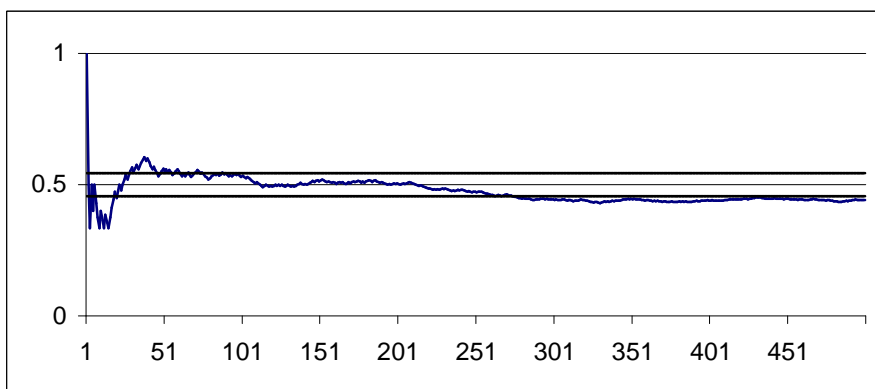
qui affecte à p la valeur $\left\{ \begin{array}{l} 0,44 \text{ une fois sur quatre} \\ 0,5 \text{ une fois sur deux} \\ 0,56 \text{ une fois sur quatre} \end{array} \right. \cdot (p = 0,5 + (-1)^{\text{int}(\text{rand}+0,5)} \times 0,06 \times \text{int}(\text{rand}+0,5))$

CASIO Graph 25 → 100	T.I. 80 - 82 - 83	T.I. 89 - 92
ViewWindow 0,500,100,0.4,0.6,0.1↵ Graph Y= 0.544↵ Graph Y= 0.456↵ 0.5+(-1)^(int(Ran#+0.5))×0.06× int(Ran#+0.5)→ P↵ 0 → N ↵ For 1 → I To 500 ↵ Int(Ran#+P) + N →N↵ Plot I , N ÷ I ↵ Next	:FnOff :ClrDraw :PlotsOff :0 → Xmin :500 → Xmax :100 → Xscl :0.4 → Ymin :0.6 → Ymax :0.1 → Yscl :DrawF 0.544 : DrawF 0.456 :0.5+(-1)^(int(rand+0.5))×0.06× int(rand+0.5)→ P :0 → N :For (I , 1 , 500) :int(rand + P) + N → N :Pt-On (I , N / I) :End	:FnOff :ClrDraw :PlotsOff :0 → xmin :500 → xmax :100 → xscl :0.4 → ymin :0.6 → ymax :0.1 → yscl :DrawFunc 0.544 : DrawFunc 0.456 : 0.5+(-1)^(int(rand()+0.5))*0.06* int(rand()+0.5)→ p :0 → n :For i , 1 , 500 :int(rand()+ p) + n → n :PtOn i , n/i :EndFor

Règle du test :

⇒ Si, après 500 lancers, la fréquence de pile est dans la bande horizontale, la pièce est ADMISE (jugée non truquée).

⇒ Si, après 500 lancers, la fréquence de pile est en dehors de la bande horizontale, la pièce est REFUSEE (jugée truquée).



Il peut se produire deux cas où le test se trompe : lorsqu'une mauvaise pièce est acceptée ou une bonne pièce rejetée.

II) Adéquation avec la loi équirépartie

On veut dans cette partie comparer les résultats observés avec les valeurs théoriques attendues, ici la loi équirépartie sur les apparitions de chacune des faces.

On lance 100 fois une pièce et obtient 55 fois Face et 45 fois Pile. Pour mesurer l'écart entre la distribution observée des fréquences f_i et celle des probabilités théoriques p_i (ici $\frac{1}{2}$), on calcule le nombre

$$d^2 = \sum_{i=1}^2 \frac{(f_i - p_i)^2}{p_i}. \text{ Si les valeurs théoriques et observées sont proches les unes des autres, } d^2 \text{ sera petit et}$$

on considérera qu'il y a adéquation entre l'expérience et la loi de probabilité équirépartie. Le problème est de savoir à partir de quelle borne, les valeurs de d^2 peuvent être considérées comme assez petites pour que le test soit jugé positif.

On réalise n fois une simulation de 100 lancers de pièces. Les résultats sont rassemblés dans le tableau suivant :

n	100	500	1000	2000
minimum	0	0	0	0
q_1	0,001 6	0,001 6	0,001 6	0,001 6
Me	0,006 4	0,003 6	0,003 6	0,003 6
q_3	0,015 7	0,014 4	0,014 4	0,014 4
d_9	0,032 4	0,025 6	0,025 6	0,025 6
maximum	0,129 6	0,211 6	0,078 4	0,221 6

Lorsque n est "assez grand", le neuvième décile d'une simulation est constant et indépendant des séries. On convient alors qu'il n'y a pas d'adéquation entre l'observation et la théorie, au risque de 10% si $d^2 > d_9$. Plus précisément, si la théorie est vraie, il y aura 10 chances sur 100 de refuser une adéquation valide.

Pour une connaissance plus précise des tests, on peut considérer la variable $D^2 = n.d^2$ qui permet le Test du χ^2 (Khi-Deux). Ce test est basé sur le calcul des distances entre les valeurs observées et les valeurs théoriques. Il est utilisable uniquement pour les lois discrètes. Si la loi est continue, on se ramène à une loi discrète en regroupant les données en classes de la forme $[x_i, x_{i+1}]$. Dans chaque classe, on doit obtenir un nombre théorique $np_i \geq 5$; l'intervalle de cette classe n'étant pas forcément régulier.

Pour mesurer ces écarts, on considère donc la variable aléatoire $D^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$

On montre que l'espérance mathématique de D^2 est indépendante de n et des p_i : $E(D^2) = k - 1$. De plus, on montre que la loi de D^2 ne dépend pas de n dès que n est assez grand ; en pratique pour $n \geq 30$, mais ne dépend que du nombre k d'issues. Cette loi est proche de celle d'une variable unique notée χ^2 dont les vingtiles v_α sont donnés par la table ci-après. lorsque la loi théorique est connue, $E(D^2) = k - 1$ est appelé "nombre de degrés de liberté" de cette variable (par exemple, la connaissance d'une loi sur un dé à 6 faces ne dépend que des 5 premières données, la 6^{ème} étant alors connue).

On convient, pour une expérience aléatoire, de rejeter la "régularité" de façon significative au seuil de 10% (ou de 5%) lorsque la valeur observée de D^2 dépasse $v_{18} = d_9$ (ou v_{19} pour 5%) :

Soit x la valeur observée de D^2 ,

si $x < v_{19}$, au seuil de 5%, on ne peut pas rejeter l'hypothèse initiale sur le modèle

si $x > d_9$, au seuil de 10%, on décide de rejeter l'hypothèse de régularité

$k - 1$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$v_{18}=d_9$	2,71	4,61	6,25	7,78	9,24	10,64	12,02	13,36	14,68	15,99	17,27	18,55	19,81	21,06	22,31
v_{19}	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51	16,92	18,31	19,67	21,03	22,36	23,68	24,00

$k - 1$	16	17	18	19	20	25	30	40	50	60	70	80	90	100
$v_{18}=d_9$	23,54	24,77	25,99	27,20	28,41	34,38	40,26	51,80	63,17	74,40	85,53	96,58	107,57	118,50
v_{19}	26,30	27,59	28,87	30,14	31,41	37,65	43,77	55,76	67,50	79,08	90,53	101,88	113,15	124,34