Peut-on simuler soi-même le hasard ?

IREM de Grenoble
Extrait de « Pratiques de la statistique. Expérimenter, modéliser, simuler. »
(Vuibert, 2007) .

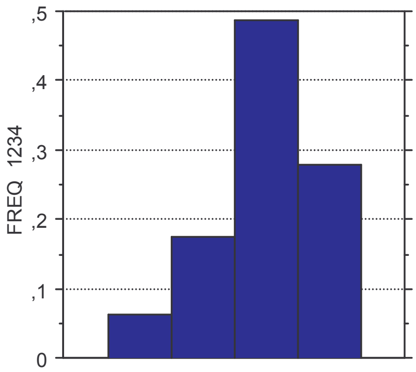
Sommes nous capables de choisir des chiffres au hasard ?
Voici une petite expérience et des premiers éléments pour les interpréter.

I- Choix du 3

On a demandé à des professeurs de mathématiques, réunis par groupe d'une centaine de personnes lors de journées de formation « interacadémiques » de répondre rapidement à la question suivante :

Choisissez au hasard un des nombres 1,2, 3 ou 4.

Les réponses sont indépendantes, et si chacun a choisi au hasard, on a un échantillon de n nombres au hasard dans {1,2,3,4}. Ici, n=424. On trouvera les résultats observés dans le tableau 1. Le nombre totale de réponses étant 424, le « nombre théorique » de réponses attendues est 106 pour chacune des 4 nombres possibles.



nombre	1	2	3	4	total
effectif	26	74	206	118	424
POURCENTAGE	6.1	17.5	48.6	27.8	100

Tableau 1 : choix de nombres dans {1,2,3,4}.

Comment juger de l'écart entre le vecteur des réponses (n_1,..,n_4) et le vecteur des réponses théoriques (106,106,106,106) ? Une des mesures possibles de cet écart est :

$$d = \sqrt{\sum_{i=1}^{4} (n_i - 106)^2}$$

On trouvera figure 1 un l'histogramme des fréquences cumulées construit à partir de 10 000 valeurs de d^2, chacune d'elle étant calculée sur un échantillon simulé de 424 nombres au hasard dans {1,2,3,4}.

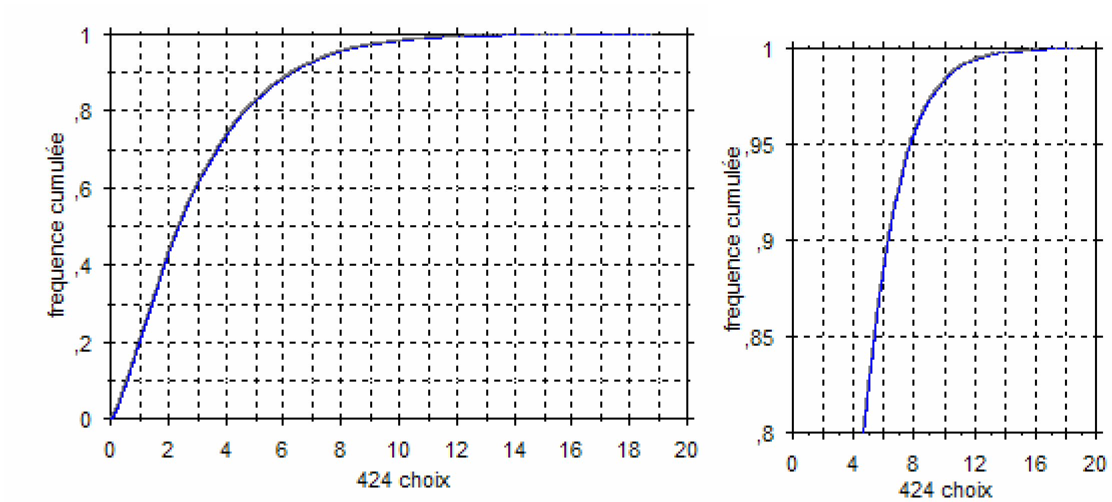


Figure 1 : histogramme des fréquences cumulées de 10 000 valeurs de d^2 pour des échantillons 424 nombres choisis au hasard dans {1,2,3,4}. On peut lire à gauche qu'environ, 75% des données sont inférieures à 4. Sur le zoom à droite, on voit qu'environ 95% des données sont inférieures ou égales à 8. Toutes les valeurs observées sont inférieures à 19.

La valeur observée pour d^2 est :

$$d^2=((26-106)^2+(74-106)^2+(206-106)^2+(118-106)^2)/106\approx 166.$$

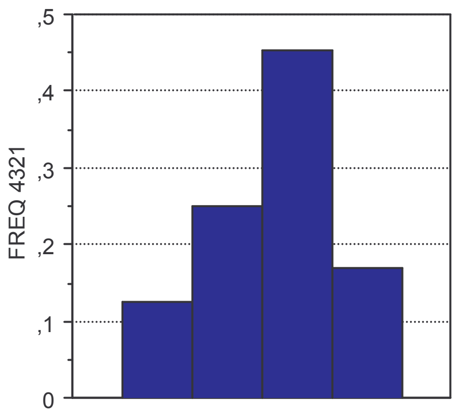
On voit que l'ordre de grandeur d^2 n'est pas le même que celui des 10 000 valeurs simulées : on rejette l'hypothèse que les 424 nombres soient choisis au hasard.

On peut montrer ici que si l'échantillon des nombres choisis au hasard est de grande taille (supérieure à 100 convient), la loi de la variable D^2 dont d^2 est une réalisation est bien approchée par ce qu'on appelle une loi du Khi-deux à 3 degrés de liberté : en particulier, la figure 1, déjà trop imprécise pour que les fluctuations d'échantillonnage se voient à l'œil nu, reste aussi en apparence la même si on remplace 424 par tout autre nombre supérieur à 100.Pour une telle loi, la probabilité que $D^2>16.3$ est inférieure à 0.001. On peut dire ici qu'au risque 0.001, on rejette l'hypothèse que les données sont un échantillon d'une loi uniforme.

Nous peut se demander si c'est le 3 qui est souvent choisi, ou le nombre en troisième position de la liste ? Pour cela, nous avons posé la question suivante :

Choisissez au hasard un des nombres 4,3,2 ou 1.

Les résultats obtenus sont consignés dans le tableau 2 ; on observe que la distribution des fréquences est sensiblement distincte de celle du tableau 1. On peut montrer, mais nous ne le ferons pas ici, qu'on ne peut pas mélanger les deux échantillons (au risque 0.05, on ne peut pas considérer que ce sont des échantillons d'une même loi).



nombre	1	2	3	4	total
effectif	37	74	134	50	295
pourcentage	12.5	25.1	45.4	16.9	100

Tableau 2 : choix de nombres dans {4,3,2,1}.

La valeur observée pour d^2 est ici

$$d^2_{obs} = ((37\text{-}m)^2+(74\text{-}m)^2+(134\text{-}m)^2+(50\text{-}m)^2)/m \approx 75 \quad \text{où } m=295/4.$$

On peut donc aussi dire qu'au risque 0.001, on rejette l'hypothèse que ces données sont un échantillon d'une loi uniforme.

II . Choisir le 7.

On a demandé à 362 personnes (en majorité des élèves de lycée, ou de stagiaires IUFM) de choisir au hasard un chiffre entre 1 et 9. Le tableau 2 présente les résultats obtenus, et la figure 2 en donne un histogramme.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**TOTAL**
EFFE	11	30	41	38	45	32	92	48	35	362
%	3	8.3	11.3	10.5	12.4	8.8	25.4	13.3	9.9	100

Tableau 2 : Choix de chiffres au hasard dans 1,2,…9.

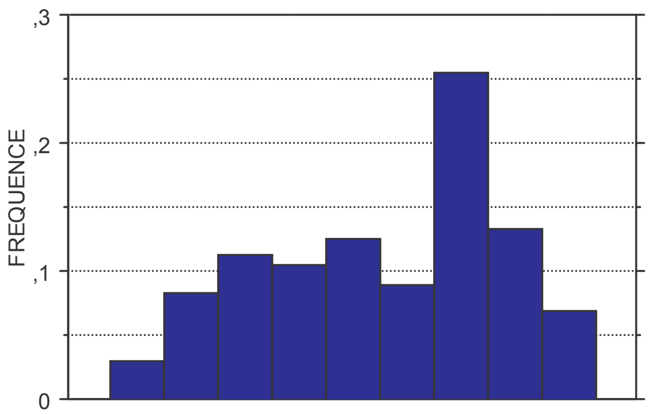


Figure 2 : Choix de chiffres au hasard dans 1,2,…9.

Il semblerait là aussi que le choix au hasard doive être remis en question.

Nous allons procéder comme précédemment, et pour mesurer globalement sur ces n observations les écarts à $n/9$, nous allons nous baser sur le valeur de l'écart entre les nombres théoriques attendus ($n/9$ pour chaque chiffre) et les nombres $n_1,..,n_9$ observés ; la mesure de l'écart choisie est :

$$d^2 = \sum_{i=1}^{9} \frac{(n_i - n/9)^2}{n/9} = \frac{9}{n}\sum_i (n_i - n/9)^2$$

On trouvera figure 3 l'histogramme des fréquences cumulées construit à partir de 10 000 valeurs de d^2, chacune d'elle étant calculée sur un échantillon simulé de 362 chiffres au hasard dans $\{1,2,\ldots,9\}$.

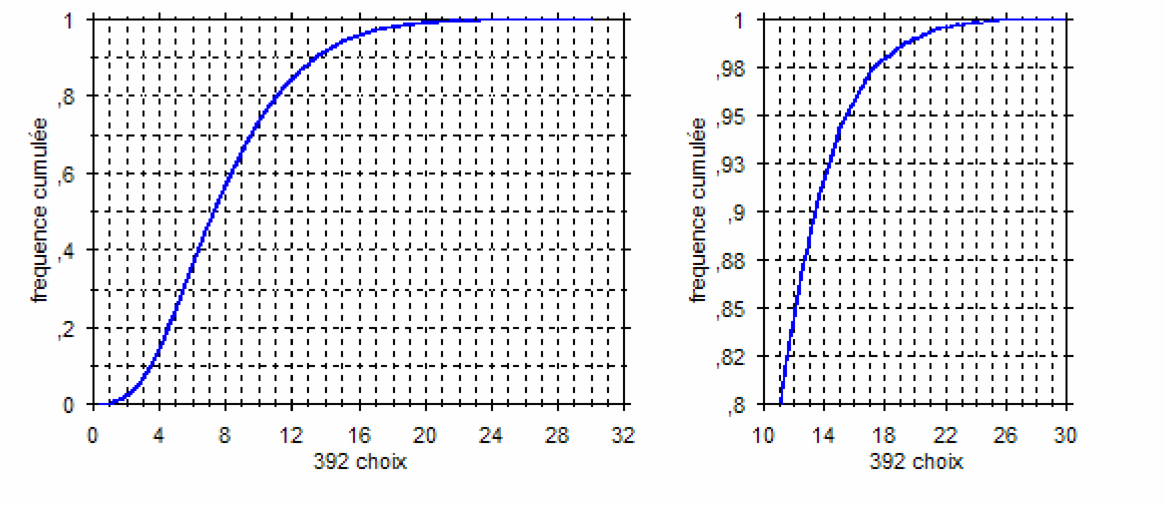


Figure 3 : histogramme des fréquences cumulées de 10 000 valeurs de d^2 pour des échantillons 392 nombres choisis au hasard dans $\{1,2\ldots,9\}$. On peut lire à gauche qu'environ, 60% des données sont inférieures à 8. Sur le zoom, à droite, on lit qu'environ 95% des données sont inférieures ou égales à 16. Par ailleurs, la plus grande valeur observée est 30.5.

La valeur observée pour d^2 est :

$$d^2=((11-r)^2+(30-r)^2+\ldots+(35-r)^2)/r \approx 88 \qquad \text{où } r=392/9.$$

On voit que l'ordre de grandeur de la valeur d^2 n'est pas le même que celle des 10 000 valeurs simulées : on rejette l'hypothèse que les 392 nombres constituent un échantillon de nombres au hasard.

On peut montrer ici que si un échantillon de nombres choisis au hasard est de grande taille (supérieure à 100), la loi de la variable D^2 est bien approchée par ce qu'on appelle une loi du Khi-deux à 8 degrés de liberté. La probabilité que $D^2 > 28$ est inférieure à 0.001. On peut dire ici qu'au risque 0.001, on rejette l'hypothèse que les données sont un échantillon de la loi de probabilité équirépartie sur $\{1,..,9\}$.

La forte présence du 3 dans la première étude et du 7 dans la seconde n'est pas spécifique au public testé : il parait qu'il en est toujours ainsi… en France.