MATH & MEDIA :
Et si on votait en tirant au hasard, à ‘pile ou face’ ?

Notre attention a été attirée par un article de Libération du 10 décembre 200, relatif aux élections américaines, et en particulier à celles de l’état de Floride. On y faisait référence à une étude de Bernard PARIZOT, astrophysicien au CNRS et amateur de statistiques, où celui affirmait (au moment où l’écart entre Bush et Gore n’était que de 537 voix pour environ 6 millions de votants) une certain nombre de propositions a priori étonnantes., dont celle ci : « *Un simple tirage à pile ou face entre les deux candidats aurait sans doute donné un écart beaucoup plus important* ».
Nous avons voulu en savoir un peu plus, et avons consulté le site < www.cyberhumanisme.org > où B. PARIZOT avait rédigé ses conclusions.
Nous avons pu y lire ceci :
(…) les écarts de voix entre les deux principaux candidats sont beaucoup plus faibles que si on tirait les votes à pile ou face.
(…) là où l'approche mathématique devient intéressante, c'est qu'elle permet de calculer précisément l'ordre de grandeur des fluctuations statistiques en fonction du nombre de pièces lancées [il s’agit là du tirage à pile ou face] *: les fluctuations statistiques sont de type gaussien, avec une largeur relative typique en racine de 1/N, où N est le nombre de lancers.*
(…) Prenons maintenant le cas du fameux Etat de la Floride, cause de tous les problèmes. Il y a paraît-il 6 millions d'électeurs. Par le simple fait du hasard, c'est-à-dire dans le cas où aucune opinion ne guiderait les votes, on devrait donc obtenir un écart de voix de l'ordre de racine carrée de 6 millions, c'est-à-dire environ 2500 voix

Cette dernière phrase pourrait être interprétée comme « L’espérance de l’écart est de l’ordre de 2 500 voix ». Même ainsi, elle est ambiguë : est-ce l’écart entre le nombre de voix observées et 3 000 000, ou l’écart entre le nombre de voix de chacun des deux candidats ? Par la suite, nous privilégions la première interprétation, sans être sûrs que ce soit la bonne…

Posons le problème de façon plus claire : on joue n fois à ‘pile ou face’ (avec une pièce équilibrée). L’espérance du nombre de ‘PILE’ est n/2. Appelons D la variable aléatoire correspondant à l’écart (en valeur absolue) entre le nombre de ‘PILE’ observé et n/2. **Il s’agit de calculer l’espérance E(D).** Remarquons tout de suite que si on cherchait à déterminer l’espérance de l’écart entre le nombre de ‘PILE’ et de ‘Face’, on trouverait 2*E(D).

Le point du vue de l’analyste combinatoire

On fera l’hypothèse que n est pair, pour que n/2 soit entier. Le nombre de ‘PILE’, X, suit une loi binomiale B(n ; ½). Posons $D = |X - n/2|$. Pour tout k (entier) compris entre 0 et n/2 on a :

$$p(D = k) = p(X = n/2 - k) + p(X = n/2 - k) = \frac{C_n^{n/2+k}}{2^{n-1}}.$$

D'où $E(D) = \frac{1}{2^{n-1}} \sum_{k=1}^{n/2} k.C_n^{n/2+k}$

Et par un astucieux calcul (si l'on sait bien manier la combinatoire), on trouve finalement que $E(D) = \frac{n.C_n^{n/2}}{2^{n+1}}$

L'inconvénient est qu'on ne sait pas calculer effectivement des combinaisons portant sur de si grandes valeurs (n'oublions pas que, dans notre exemple, n = 6 000 000). L'analyste va donc chercher la limite de cette expression lorsque $n \to \infty$. Et cela en utilisant les intégrales de Wallis $\omega_k = \int_0^{\pi/2} \sin^k(x).dx$, sachant que $\omega_n = \frac{\pi}{2} \cdot \frac{C_n^{n/2}}{2^n}$. On trouve donc, in fine, que E(D) est de l'ordre de grandeur de $\sqrt{\frac{n}{2\pi}}$ lorsque n est 'infiniment grand'.

A propos des intégrales de Wallis, consultez l'excellent ouvrage de B. et A. PARZYSZ *__Fonctions d'une variable__* dans la collection TD chez Dunod. Le calcul ci-dessus a d'ailleurs été fait par Bernard PARZYSZ, que nous remercions.

Le point de vue du probabilitste-statisticien

Le statisticien, lui, sait que la loi binomiale B(n ; ½) de X (nombre de 'PILE') est approchée, dès que n est suffisamment grand, par une loi normale N(n/2 ; √n/2).

La fonction de répartition de D est définie par $F_D(x) = p(|X - n/2| < x)$ = p (n/2 x < X < n/2 + x) pour $x \geq 0$.

Soit $F_D(x) = F_X(n/2 + x) - F_X(n/2 - x)$. En dérivant, il vient :

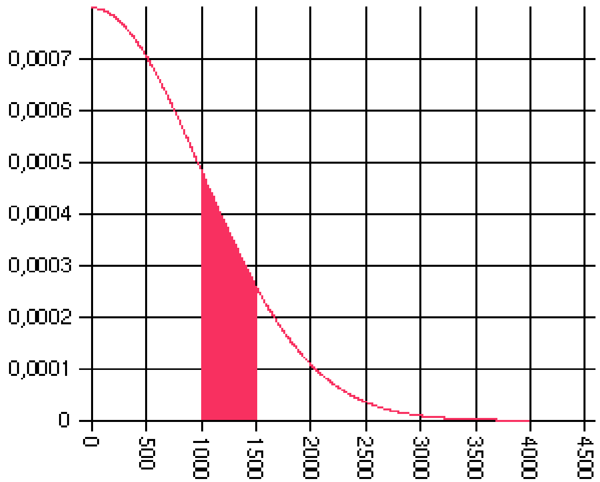
$f_D(x) = f_X(n/2 + x) - f_X(n/2 - x)$, où f_X est la densité de X, dont on disait qu'on l'avait assimilée à une loi normale.

On trouve donc $f_D(x) = \frac{2\sqrt{2}}{\sqrt{n\pi}} \cdot e^{\frac{-2x^2}{n}}$

.D'où $E(D) = \int_0^{\infty} x.f_D(x).dx = \frac{\sqrt{n}}{\sqrt{2\pi}} \cdot \left[-e^{\frac{-2x^2}{n}} \right]_0^{\infty}$.

Ce qui, in fine, redonne bien $E(D) = \sqrt{\frac{n}{2\pi}}$. Merci à Daniel Vagost et Franck Gaüzère, du département S.T.I.D. de l'I.U.T. de Metz, d'avoir ment à bien ce calcul.

L'image ci-dessous est la densité de 2D (différence en valeur absolue entre le nombre d PILE et le nombre de FACE (pour n = 1 000 000).
Nous l'avons trouvée sur le site personnel de Michel Brissaud, qui a lui aussi traité ce problème de l'écart des voix (avec l'aide d'un tableur). Nous l'en remercions et nous vous conseillons de consulter sa page personnelle :
<http://www.perso.wanadoo.fr/michel.brissaud>



Le point de vue du simulateur (!)
Le simulateur est l'auteur de ces lignes : il ne sait ni calculer des intégrales (fussent-elles de Wallis, de Futuna ou d'ailleurs), ni manier correctement la combinatoire. Il a donc réalisé un petit programme pour simuler 6 000 000 de tirages à 'pile ou face', et regardé l'écart (en valeur absolue) entre le nombre de 'PILE' et 3 000 000. Et il a répété 500 fois cette expérience aléatoire.
Voici l'algorithme :

```
nbvotants=6000000 ; nbA=0, écart=0, totalécart=0;
écartmax=0, écartmin=nbvotants; écartmoy=0;
entrer nbrepet; (équivalent d'un input)
for(j=0;j<nbrepet;j=j+1)
    { nbA=0; écart=0;
    for (i=0;i<nbvotants;i=i+1) if (random()<0.5) nbA=nbA+1;
    écart=abs(nbvotants/2-nbA); totalécart=totalécart+écart;
    if (écart>écartmax) écartmax=écart;
    if (écart<écartmin) écartmin=écart;
    } (fin de la boucle for)
écartmoy=totalécart/nbrepet;
print("Sur "+nbrepet+" expériences, écart moyen =
"+écartmoy);
print("Maximum de l'écart : "+écartmax+" ; minimum de
l'écart : "+écartmin);
```

Le langage de cet algorithme est purement imaginaire, mais il s'inspire du C ou du Java. Au bout d'un certain temps (500 fois 6 000 000, ça fait quand même 3 milliards de tirages), le résultat s'affiche :

```
Sur 500 expériences, écart moyen = 958,2
Maximum de l'écart : 4066 ; minimum de l'écart : 9
```

Et on est satisfait, car cela correspond au résultat théorique !

Si on avait recherché, par simulation, la fréquence d'un écart supérieur à $\sqrt{n}/2$ (ce qui correspond aux dires de B. PARIZOT, voir ci-dessous), on aurait trouvé environ 1/3 : ce n'est pas un événement si rare que ça.

Mais… car il y a un mais !

Bernard PARIZOT a écrit « *Par le simple fait du hasard, c'est-à-dire dans le cas où aucune opinion ne guiderait les votes, on devrait donc obtenir un écart de voix de l'ordre de racine carrée de 6 millions, c'est-à-dire environ 2500 voix* ».
Ce qui ne correspond ni à ce qu'on a calculé, ni à ce qu'on a simulé.
On a certainement mal interprété son texte : il parlait certainement de l'écart entre les deux candidats, et non de l'écart entre le nombre de voix de l'un des deux et 3 000 000. Il faut donc doubler tous les résultats trouvés auparavant.

Mais cela ne donne toujours qu'un espérance de $\sqrt{\frac{2n}{\pi}}$, soit environ 1 954 voix.
L'explication tient peut-être dans cette phrase : « *les fluctuations statistiques sont de type gaussien, avec une largeur relative typique en racine de 1/N* ». Ce qui correspond à la « plage de normalité » dont il est question dans le programme de seconde à propos des fluctuations d'échantillonnage : quand ou joue à 'pile ou face', on a en effet 95% de chances d'observer, dans l'échantillon tiré, que la proportion de 'PILE' est dans l'intervalle

$$\left]\frac{1}{2}-\frac{1}{\sqrt{n}}\,;\frac{1}{2}+\frac{1}{\sqrt{n}}\right[.$$

L'auteur pose inconsciemment le problème de la vulgarisation scientifique, qui oblige parfois à une trop grande simplification du propos pour qu'il puisse être compris par tous. Mais le cœur du message est présent : l'écart est dans l'ordre de grandeur de $\sqrt{n}$…

Jacques VERDIER

P.S. Dans un message électronique du 27/01/01, Michel BRISSAUD m'écrivait ceci : « *En conclusion, c'est très compliqué, et je continue à penser qu'un approche correcte de la fluctuation d'échantillonnage est impossible en classe de seconde : il faut rester à un niveau très élémentaire.* »