Un peu de probabilités et de statistiques

Sandrine Caruso

29 mars 2009

1 Les coefficients binomiaux

Exercice 1. Une urne contient 6 boules numérotées de 1 à 6. On en tire 2 au hasard. Combien y a-t-il de tirages possibles (sans tenir compte de l'ordre dans lequel on les a tirées) ?

Plus généralement, si l'urne contient n boules et qu'on en tire k, montrer qu'il y a $\frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$ tirages possibles. Comment exprimer ce nombre à l'aide de factorielles d'entiers ?

On note ce nombre $\binom{n}{k}$. Les $\binom{n}{k}$ sont appelés des *coefficients binomiaux*, en raison de la formule suivante.

Formule 1 (Binôme de Newton). *Soient* $a, b \in \mathbb{R}$ *et* $n \in \mathbb{N}$.

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = \binom{n}{0} a^0 b^{n-0} + \binom{n}{1} a^1 b^{n-1} + \cdots + \binom{n}{n} a^n b^0 \tag{1}$$

Exercice 2. Montre les formules suivantes :

- $$\binom{n}{n-k} = \binom{n}{k} \tag{2}$$
- $$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \tag{3}$$
- $$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{4}$$
- $$\sum_{k=0}^{n} k \binom{n}{k} a^k b^{n-k} = na(a+b)^{n-1} \tag{5}$$

Comment la formule (3) permet-elle de construire le *triangle de Pascal* ? (Si tu ne sais pas ce qu'est le triangle de Pascal, une recherche sur internet saura te renseigner...) Comment utiliser la formule (1) facilement à l'aide du triangle de Pascal ?

2 Notions de probabilités

2.1 Loi de Bernouilli

Si l'on lance une pièce de monnaie, on a deux choix possibles : soit elle tombe sur « pile », soit elle tombe sur « face »[1]. En général, une pièce n'est pas parfaitement équilibrée, et il se peut qu'elle ait plus de chances de tomber d'un côté que de l'autre. Notons p la probabilité que la pièce ait de tomber sur « pile ». La probabilité qu'elle a de tomber sur « face » est alors égale à $1 - p$. On associe à la pièce une *variable aléatoire*, que l'on note X, qui vaut 1 si la pièce tombe sur « pile » et 0 si la pièce tombe sur « face ».

On dit que X suit *une loi de Bernouilli de paramètre* p.

Exercice 3. Que vaut X en moyenne ? En termes probabilistes, la *moyenne* est appelée *espérance* (si tu gagnes 1 euro lorsque la pièce tombe sur pile et 0 euro lorsqu'elle tombe sur face, l'espérance de X est ce que tu peux espérer gagner en moyenne).

2.2 Univers, événements

On suppose que l'on a une expérience aléatoire (lancer de pièce, de dés, piocher des boules dans une urne, etc...) qui ne peut produire qu'un nombre fini de résultats arrivant chacun avec une certaine probabilité. On note Ω leur ensemble, c'est ce qu'on appelle l'*univers*. Si $\omega \in \Omega$ est un résultat possible, notons $p(\omega)$ sa probabilité.

Un *événement* E est une partie de Ω. Sa probabilité est la somme des probabilités des résultats qu'il contient, on la note $\mathbb{P}(E)$.

Exemple 4. Considérons l'expérience consistant à lancer une pièce deux fois. L'univers est formé de 4 résultats : (pile, pile), (pile, face), (face, pile) et (face, face). Leurs probabilités respectives sont p^2, $p(1 - p)$, $(1 - p)p$ et $(1 - p)^2$. Un exemple d'événement est le sous-ensemble {(pile, pile), (pile, face)} : c'est l'événement « la premier lancé a donné pile ». Sa probabilité est $p^2 + p(1 - p) = p$... Ce qui est plutôt rassurant !

Exercice 5. On lance la pièce N fois. Combien y a-t-il de résultats possibles ? Combien en compte l'événement « la première pièce est tombé sur pile » ?

2.3 Variable aléatoire, espérance, variance, écart-type

Maintenant, on attribue à chaque résultat une valeur (un « gain »), par exemple, dans le cas d'une pièce avec la loi de Bernouilli, 1 pour pile et 0 pour face. Ou la somme des résultats des dés dans le cas d'un jet de dés...

C'est ce qu'on appelle une *variable aléatoire* X. On note $X(\omega)$ le gain attribué au résultat ω. On peut faire des opérations sur les variables aléatoires. Par exemple, la variable aléatoire X^2 est celle pour laquelle le gain de ω est $X(\omega)^2$.

Si a est un nombre réel, il est courant de s'intéresser aux événements « $X = a$ »(resp. « $X \geqslant a$ », « $X \leqslant a$ »). Il s'agit de l'ensemble des ω tels que $X(\omega) = a$ (resp $X(\omega) \geqslant a$, $X(\omega) \leqslant a$). On note $\mathbb{P}(X = a)$ (resp. $\mathbb{P}(X \geqslant a)$, $\mathbb{P}(X \leqslant a)$) sa probabilité.

[1] En théorie, elle pourrait aussi tomber sur « tranche », mais on va négliger cette possibilité.

L'espérance de X est donnée par la formule

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p(\omega) X(\omega)$$

où la symbole $\sum_{\omega\in\Omega}$ signifie que l'on additionne toutes les quantités $p(\omega)X(\omega)$, pour tous les résultats ω de Ω.

Exercice 6. Donne une formule pour l'espérance de X^2.

Quelle est l'espérance d'une variable aléatoire constante (c'est-à-dire telle que toutes les possibilités donnent le même gain) ?

Si X et Y sont deux variables aléatoires sur le même univers (avec les mêmes probabilités), quelle est l'espérance de $X + Y$ en fonction de celles de X et de Y ?

Si E est un événement, on note $\mathbb{1}_E$ la variable aléatoire telle que $\mathbb{1}_E(\omega)$ est égal à 1 si $\omega \in E$ et 0 sinon. Calcule son espérance.

Exercice 7. Que donne la formule de l'espérance dans le cas du pile ou face ? Est-ce que tu retrouves le résultat trouvé à l'exercice 3 ?

Quelle est l'espérance du lancer d'un dé équilibré à 6 faces ? De deux dés équilibrés à 6 faces ?

Connaître l'espérance d'une expérience aléatoire donne certaines informations sur cette expérience... Pourtant, des expériences très différentes peuvent donner la même espérance ! (Un élève qui a 8 en français et 18 en maths, et un qui a 13 en français et 13 en maths, ont la même moyenne...) Il peut être intéressant de mesurer l'écart entre les résultats : est-ce que l'expérience donne toujours des résultats très proches les uns des autres, ou au contraire des résultats très variés ?

Pour cela, les probabilistes ont inventé un autre outil, que l'on appelle la *variance*, qui mesure l'écart à la moyenne. Si X est une variable aléatoire, sa variance est

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

Remarque que ce nombre est toujours positif : la variable aléatoire $(X - \mathbb{E}(X))^2$ est toujours positive, donc sa moyenne est également toujours positive. On appelle *écart-type* la racine carrée de ce nombre : $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

Exercice 8. Montrer que, si X est une variable aléatoire, $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Exercice 9. Quelles sont la variance et l'écart-type de la loi de Bernouilli de paramètre p ?

2.4 Loi binomiale

On est toujours en possession d'une pièce qui tombe sur pile avec une probabilité p et sur face avec une probabilité $1 - p$ que l'on lance plusieurs fois d'affilée, et on compte le nombre de pile et le nombre de face.

Exercice 10. On lance la pièce N fois. Quelle est la probabilité p_k que l'on obtienne k fois pile (k va de 0 à N) ? Pourquoi cette loi s'appelle-t-elle la *loi binomiale (de paramètres N et p)* ?

Vérifie que la somme $p_0 + \cdots + p_N$ des probabilités obtenues fait bien 1.

Exercice 11. Prenons l'exemple de $N = 8$ et $p = \frac{1}{2}$. Trace, sur un diagramme en bâtons, p_k en fonction de k.

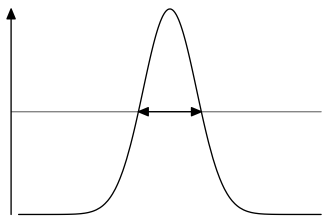
Exercice 12. Va sur le site

`http ://www.espace-sciences.org/science/images/images-maj/Perso/manipulations/planche_clous/index.htm`,

et joue autant de fois que tu veux... Comment interpréter cette expérience ? Compare le diagramme obtenu à celui que tu as tracé à l'exercice 11. Montre que cette expérience suit la loi binomiale de paramètres 8 et $\frac{1}{2}$.

Exercice 13. Quelle est l'espérance de la loi binomiale si le gain de chaque « pile » est 1 et celui de chaque « face » est 0 ? Quelles sont la variance et l'écart-type ? Essaie de donner ces résultats sous une forme simple.

Exercice 14. Trace des diagrammes comme à l'exercice 11, en choisissant une valeur de p, avec des valeurs de N de plus en plus grandes. Mesure la largeur du « pic » à la moitié de sa hauteur (comme sur le dessin ci-dessous). Cela mesure à peu près l'écart-type de l'expérience. Divise cette largeur par $\sqrt{N}$; que remarques-tu ? Essaie avec d'autres valeurs de p. Comment ce quotient dépend-il de p (tu peux tracer une courbe représentant le quotient en fonction de p) ? Essaie également de mesurer la largeur des pics à d'autres hauteurs...



3 Les sondages

Il arrive souvent que l'on cherche à estimer une loi de probabilité qu'on ne connaît pas. C'est là qu'interviennent les statistiques. Prenons l'exemple d'une élection où il y a deux candidats, que l'on va appeler « pile » et « face ». Supposons qu'un certain jour (avant l'élection), la proportion d'électeurs qui pensent voter pour « pile » est p ; la proportion d'électeurs qui pensent voter pour « face » est alors $1 - p$ (en supposant, pour simplifier, qu'il n'y a pas de vote blanc ou nul).

Le but d'un sondage est d'essayer d'estimer p. Si l'on interroge une personne au hasard (et qu'elle répond effectivement ce qu'elle pense voter), il y a une probabilité de p qu'elle réponde « pile » et une probabilité de $1 - p$ qu'elle réponde « face ».

Exercice 15. On réalise un sondage parmi N personnes. Le nombre de personnes votant pour « pile » parmi ces N suit une certaine loi de probabilités rencontrée plus haut... Laquelle ?

Comment interpréter p sur les diagrammes dessinés à l'exercice 14 ?

Le principe d'un sondage est d'interroger un nombre suffisant de personnes afin de se faire une idée de p.

On note $p(N)$ la proportion de personnes, parmi les N sondés, qui pensent voter « pile » ; $p(N)$ donne une approximation de p. On se propose d'évaluer l'incertitude du sondage, c'est-à-dire la différence entre p et $p(N)$. Si l'on a pas de chances, cette différence pourrait être grande (par exemple s'il n'y a qu'une personne qui veut voter « pile » et que l'on tombe par hasard N fois sur celle-ci). Remarquons également qu'il y a peu de chances que $p(N)$ soit *exactement* égal à p, mais on veut essayer d'évaluer la probabilité que $|p(N) - p|$ soit plus petit qu'un certain nombre ε strictement positif.

Exercice 16. Comment interpréter sur les graphiques de l'exercice 14 l'ordre de grandeur de $\mathbb{P}(|p(N) - p| \leqslant \varepsilon)$?

Remarque que, jusqu'à présent, on supposait que p était connu, et on étudiait la probabilité que $p(N)$ prenne telle ou telle valeur. Dans un sondage, on connaît $p(N)$ et c'est p que l'on essaie d'évaluer.

Exercice 17. Dans les diagrammes de la partie précédente, tu avais tracé la probabilité d'avoir k fois « pile » (c'est-à-dire $\mathbb{P}(p(N) = \frac{k}{N})$) en fonction de k, connaissant p. Cette fois, trace $\mathbb{P}(p(N) = \frac{k}{N})$ en fonction de p avec k fixé (c'est le nombre de « pile » donné par le sondage), en choisissant $N = 500$ et $k = 255$. Tu pourras utiliser $\binom{500}{255} \simeq 1{,}057 \times 10^{149}$.

Comment interprètes-tu cette courbe ? Que penser d'un sondage réalisé parmi 500 personnes, et qui indique 51% de suffrages pour « pile » ? Peut-on en déduire la victoire probable de « pile » ?