Quelques surprises des probabilités
27/4/08,
Lycée Germaine de Staël, Montluçon

*

1. Les deux enveloppes
2. Dés de fantaisie
3. Anniversaires
4. La nouvelle secrétaire
5. Exponentielle
6. Le bus qui n'arrive pas.

*G. Letac, Laboratoire de Probabilités et Statistique,
Université Paul Sabatier, Toulouse.

L'enveloppe de Karine Pompon Heureux candidat vous êtes admis à participer à notre jeu télévisé : Karine Pompon l'animatrice vous tend deux enveloppes contenant de l'argent : l'une contient une somme d'argent a l'autre une somme d'argent b avec $a < b$ sous forme d'un chèque afin de ne pas se laisser guider par l'épaisseur de l'enveloppe. Attention, vous ne connaissez si a ni b. Vous en choisissez une, vous regardez son contenu, appelons le C. Comme vous ne connaissez ni a ni b vous ne savez pas si $C = a$ ou si $C = b$. Maintenant, Karine Pompon vous offre

- ou bien de garder l'enveloppe que vous avez choisie et de rentrer chez vous avec la somme qu'elle contient ;
- ou bien de prendre l'autre enveloppe et de partir avec.

J'ai été interloqué quand un collègue m'a demandé ce que je ferais pour avoir de meilleures chances de partir avec la forte somme b : il me paraissait évident que mes chances ne dépasseraient jamais 50% quoi que je fasse. Erreur, erreur. On peut faire beaucoup mieux si on dispose d'une simple table de nombres au hasard (c'est à dire d'un moyen de tirer au sort une variable aléatoire uniforme U sur $(0,1)$). Rappelons que si $0 < u < 1$ alors $\Pr(U < u) = u$.

Voici ce que je fais : je transforme d'abord ma variable aléatoire uniforme U en une variable aléatoire exponentielle $X = -\log U$. Je dis que X est exponentielle pour la raison suivante si $x > 0$ on a

$$\begin{aligned} \Pr(X > x) &= \Pr(-\log U > x) \\ &= \Pr(\log U < -x) \\ &= \Pr(U < e^{-x}) = e^{-x}. \end{aligned}$$

Avantage, la loi de X est répartie sur toute la demi droite des nombres réels positifs.

Maintenant pendant que Miss Pompon me tarabuste en me demandant ce que je décide, je tire au sort U, j'en déduis X. A ce point je regarde si X est plus grand que C ou non.

- Ou bien $X > C$. Dans ce cas je parie que $C = a$ et je choisis donc l'autre enveloppe.
- Ou bien $X \leq C$. Dans ce cas je parie que $C = b$ et je garde la première enveloppe tirée.

Bah, c'est une procédure comme une autre, mais son succès n'est pas garanti. Certes, certes. Mais voulez vous savoir quelles sont mes chances de partir avec la grosse somme b? Réponse bien au dessus de 50% puisque c'est :

$$\frac{1}{2} + \frac{1}{2}\Pr(a < X \leq b) = \frac{1}{2} + \frac{1}{2}(e^{-a} - e^{-b}).$$

Le démontrer exige de comprendre ce qu'est l'indépendance et la disjonction d'événements. La quantité d'argent C dans l'enveloppe que j'ai tirée est une variable aléatoire dont voici la loi :

$$\text{Pr}(C = a) = \frac{1}{2}, \quad \text{Pr}(C = b) = \frac{1}{2}.$$

Comme je ne connais pas a et b je ne connais pas très bien C non plus, tant pis. Enfin j'introduis une troisième variable aléatoire V qui dépend entièrement de X et de C et définie ainsi

$$V = C \text{ si } X \leq C, \; V = a + b - C \text{ si } X > C.$$

D'après la procédure décrite ci dessus, V est bien la quantité d'argent que j'emporterai chez moi.

Voici alors le tableau des 4 possibilités, des 4 événements disjoints

$$\begin{aligned}
\{X \leq C\} \cap \{C = a\} &= \{X \leq a\} \cap \{C = a\} \\
\{X \leq C\} \cap \{C = b\} &= \{X \leq b\} \cap \{C = b\} \\
\{X > C\} \cap \{C = a\} &= \{X > a\} \cap \{C = a\} \\
\{X > C\} \cap \{C = b\} &= \{X > b\} \cap \{C = b\}.
\end{aligned}$$

	$C = a$	$C = b$
$X \leq C$	$V = a$	$V = b$
$X > C$	$V = b$	$V = a$

A chacun d'entre eux correspond une valeur de V et donc la probabilité qui nous intéresse est $\Pr(V = b)$

$$
\begin{aligned}
&\overset{(1)}{=} \Pr([\{X \le b\} \cap \{C = b\}] \cup [\{X > a\} \cap \{C = a\})] \\
&\overset{(2)}{=} \Pr[\{X \le b\} \cap \{C = b\}] + \Pr[\{X > a\} \cap \{C = a\})] \\
&\overset{(3)}{=} \Pr(X \le b)\Pr(C = b) + \Pr(X > a)\Pr(C = a) \\
&\overset{(4)}{=} \Pr(X \le b)\frac{1}{2} + \Pr(X > a)\frac{1}{2}.
\end{aligned}
$$

Pardon de cette longue chaîne d'égalités, mais chacune d'elle illustre ce que vous apprennent vos professeurs : la ligne (1) est une autre description de l'événement $V = b$. La ligne (2) illustre les événements disjoints. La ligne (3) illustre les événements indépendants. Pour terminer

$$
\begin{aligned}
&\Pr(X \le b) + \Pr(X > a) = \\
&\Pr(X \le a) + \Pr(a < X \le b)) \\
&+ \Pr(a < X \le b)) + \Pr(b < X) \\
&= 1 + \Pr(a < X \le b)).
\end{aligned}
$$

Dés de fantaisie. Si on équipe les faces de deux dés équilibrés des points $\{1, 2, 2, 3, 3, 4\}$ et $\{1, 3, 4, 5, 6, 8\}$ on obtient deux variables aléatoires indépendantes X et Y de lois respectives

$$\frac{1}{6}(\delta_1 + 2\delta_2 + 2\delta_3 + \delta_4)$$

et

$$\frac{1}{6}(\delta_1 + \delta_3 + \delta_4 + \delta_5 + \delta_6 + \delta_8)$$

alors que si les deux dés sont classiques on obtient deux va U et V indépendantes et de même loi $\frac{1}{6}(\delta_1+\delta_2+\delta_3+\delta_4+\delta_5+\delta_6)$. La chose amusante est que $X+Y$ et $U+V$ ont la même loi. Ceci ne permet toutefois pas de jouer au Monopoly avec ces dés de fantaisie, car le cas des doubles (cad le cas où on tire $U = V$) a un rôle dans le Monopoly puisque un double permet de rejouer et que 3 doubles à la suite mènent en prison. Or $\Pr(U = V) = \frac{1}{6}$ alors que $\Pr(X = Y) = \frac{1}{9}$.

Je voudrais expliquer qu'il n'y a pas d'autres dés de fantaisie possibles que ceux là, à part la solution triviale qui consiste à remplacer les X et Y ci dessus par $(X-1, Y+1)$, $(X+1, Y-1)$, $(U+1, V-1)$ conduisant à des faces sans point du tout. La raison est que si on note $A(x) = 1 + x$, $B(x) = 1 + x + x^2$ et $C(x) = 1 - x + x^2$ alors la décomposition du polynôme $\mathbb{E}(x^{U+V})$ en facteurs irréductibles réels est

$$\mathbb{E}(x^{U+V}) = \frac{x^2}{36} A(x)^2 B(x)^2 C(x)^2$$

Si X et Y sont les va associées à des dés de fantaisie alors les polynômes P et Q tels que $\frac{1}{6}P(x) = \mathbb{E}(x^X)$ et $\frac{1}{6}Q(x) = \mathbb{E}(x^Y)$ doivent satisfaire donc à trois conditions :

1. $P(x)Q(x) = x^2 A(x)^2 B(x)^2 C(x)^2$;
2. être à coefficients entiers ≥ 0;
3. satisfaire $P(1) = Q(1) = 6$.

Il existe donc deux nombres $p, q > 0$ avec $pq = 1$ et 4 entiers a, b, c, d dans $\{0, 1, 2\}$ tels que $P = px^d A^a B^b C^c$ et $Q = qx^{2-d} A^{2-a} B^{2-b} C^{2-c}$. Comme P/x^d est à coefficients entiers, en faisant $x = 0$ on voit que p est entier. Même chose pour q et donc $p = q = 1$. Ensuite

$$P(1) = 2^a 3^b = 6,$$

ce qui impose $a = b = 1$. On voit que la seule liberté qui subsiste est $c = 0, 1, 2$ et $d = 0, 1, 2$. Faire varier d conduit aux trivialités exclues plus haut. Prenant $d = 1$ on voit que $c = 1$ est le cas classique, $c = 0$ est la paire fantaisie du début, et $c = 2$ est naturellement son symétrique.

Happy birthday n personnes sont réunies. Quelle est la probabilité que au moins deux d'entre elles aient le même anniversaire ?

On formalise le problème en le simplifiant un peu : on ignore d'abord le problème du 29 février, et on postule donc que l'espace des observables est $\Omega = F^E$ où E est l'ensemble des personnes et où F est l'ensemble des $p = 365$ jours de l'année : on observe donc la fonction $f \in \Omega$ qui à chaque personne associe son anniversaire. On postule ensuite qu'on est dans le cas équiprobable. Finalement, il est plus facile de calculer la probabilité du complémentaire A^c de l'événement A "deux personnes au moins ont le même anniversaire", car c'est la probabilité que la fonction f soit injective.

C'est $\Pr(A^c) =$

$$P(A^c) = \frac{1}{365^n} 365(365-1)\cdots(365-n+1) =$$

$$\prod_{k=1}^{n-1}(1-\frac{k}{365}) = \exp\sum_{k=1}^{n-1}\log(1-\frac{k}{365}).$$

Si n n'est pas grand, une évaluation approximative de cette somme se fait en remplaçant $\log(1-x)$ par $-x$ et en utilisant la somme d'une progression arithmétique étudiée en Terminale

$$\sum_{k=1}^{n-1} k = \frac{1}{2}n(n-1) \sim n^2/2,$$

qui conduit à l'approximation $\Pr(A^c) \sim \exp(-n^2/730)$. Pour voir par exemple pour quel n on a $\Pr(A^c) \sim 1/2$ on prend $n \sim \sqrt{730 \log 2} \sim 23$.

Pour un calcul plus sérieux, on peut utiliser l'encadrement pour $0 < x < 1$:

$$-x - \frac{x^2}{2(1-x)} < \log(1-x) < -x - \frac{x^2}{2};$$

La majoration de droite se déduit du développement en série entière, celle de gauche se montre en étudiant la fonction $x + \frac{x^2}{2(1-x)} + \log(1-x)$. On a aussi besoin de la somme des premiers carrés :

$$\sum_{k=1}^{n-1} k^2 = \frac{1}{6} n(2n-1)(n-1) \sim n^3/3,$$

qui s'établit par récurrence. Si $x \leq (n-1)/365$, alors $-1/(1-x) \geq -365/(365-n+1)$.

D'où l'encadrement :

$$-\frac{n(n-1)}{2}\frac{1}{365}-\frac{n(n-1)(2n-1)}{6}\frac{1}{2\times 365^2}\frac{365}{365-n+1}$$

$$< \log \Pr(A^c) <$$

$$-\frac{n(n-1)}{2}\frac{1}{365} - \frac{n(n-1)(2n-1)}{6}\frac{1}{2\times 365^2}.$$

Par exemple, si $n = 35$ on trouve $\text{Pr}(A^c) = 0,186...$ Si 35 personnes sont réunies, la probabilité que deux d'entre elles au moins aient le même anniversaire est donc $0,813...$ Si $n = 40$ $\text{Pr}(A^c) = 0,110...$, si $n = 51$ $\text{Pr}(A^c) = 0,039...$ et si $n = 60$ $\text{Pr}(A^c) = 0,007...$

$$\begin{array}{ll} n = 23 & \text{Pr}(A) = 0,499... \\ n = 35 & \text{Pr}(A) = 0,813... \\ n = 40 & \text{Pr}(A) = 0,889... \\ n = 51 & \text{Pr}(A) = 0,960... \\ n = 60 & \text{Pr}(A) = 0,992... \end{array}$$

Le problème des secrétaires. Madame le Proviseur a besoin d'une nouvelle secrétaire : 30 candidates se présentent, il s'agit de choisir la meilleure, car elles ne se valent pas : en fait si on pouvait les examiner toutes ensembles, on leur donnerait à chacune une note entre 1 et 30, toutes distinctes. Malheureusement pour Madame le Proviseur, le chômage a disparu et en entendant successivement chacune des candidates, elle doit lui dire immédiatement si elle l'embauche ou non. Historiquement cela s'appelle le problème des secrétaires, et cela parait un peu sexiste et années 1950. Mais si vos parents ont un studio à louer et qu'ils cherchent le meilleur locataire, ils pourraient bien se trouver devant un semblable problème.

Voici donc la procédure que choisit Madame le Proviseur : elle entend sans les embaucher les $r-1$ premières candidates, et choisit la première qui entre r et 30 sera meilleure que chacune des $r-1$ premières. Si malheureusement personne n'est meilleur, cad si celle qui avait la note 30 est déjà passée parmi les $r-1$ premières, il faudra prendre la candidate numéro 30. Formellement si vous voulez, on note par

$$\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_{30})$$

une bijection quelconque de $\{1, 2, \ldots, 30\}$ dans lui même : σ_1 est la note qu'obtiendrait la première candidate à passer devant Madame le Proviseur si on avait pu les examiner toutes ensemble, σ_2 est la note qu'obtiendrait la seconde etc. Si $j \in \{1, 2, \ldots, 30\}$, notons

$$M_j = \max\{\sigma_1, \ldots, \sigma_j\},\ J = \min\{j; j \geq r \text{ et } M_j = \sigma_j\}.$$

Attention l'ensemble définissant J peut être vide et J sera non défini.

Où sont les probabilités et l'aléatoire ? Dans le fait que les σ sont chacun de probabilité égale à $1/(30!)$. Dans ces conditions, on a l'égalité

$$\Pr(\{M_j = 30\} \cap \{J = j\}) = \frac{1}{30} \times \frac{r-1}{j-1}.$$

Malgré la simplicité de cette formule, je renonce à vous la montrer rigoureusement. Elle vous donne la probabilité de choisir la meilleure secrétaire parmi les 30 lors de l'entrevue $j \geq r$. Plus intéressant est la facon dont on l'exploite : elle entraine d'abord que la probabilité de choisir la secrétaire avec la note 30 est

$$\Pr(M_J = 30) = \frac{1}{30}\left[1 + \frac{r-1}{r} + \frac{r-1}{r+1} + \ldots + \frac{r-1}{29}\right]$$

formule peu appétissante, mais dont on va donner une approximation très intéressante :

Je commence par remplacer 30 par n et appeler $P(r,n) = \Pr(M_J = n)$ car il ne faut pas perdre de vue que nous n'avons pas encore choisi r. Bref

$$P(r,n) = \frac{1}{n} \sum_{j=r}^{n} \frac{r-1}{j-1}.$$

Nous approchons $P(r,n)$ en écrivant $r = xn$ avec x fixé et $0 < x < 1$ et avec n grand :

$$P(r,n) \sim \frac{1}{n} \sum_{j=xn}^{n} \frac{xn}{j} = x \sum_{j=xn}^{n} \frac{1}{j} \sim -x \log x$$

(Détail :

$$\log \frac{n+1}{xn} = \int_{xn}^{n+1} \frac{dt}{t} < \sum_{j=xn}^{n} \frac{1}{j} < \int_{xn-1}^{n} \frac{dt}{t} = \log \frac{n}{xn-1})$$

Et maintenant, pour quel x avec $0 < x < 1$ le nombre $-x \log x$ est il le plus grand ? Réponse $x = e^{-1} = 0,36$. Que doit faire Madame le Proviseur avec 30 candidates ? Laisser passer $30 \times 0,36 \sim 11 - 1 = 10$ candidates et prendre la meilleure qui se présentera à partir de la 11 ème.

Importance de la loi exponentielle. La table de nombres au hasard était une table de réalisations de la loi uniforme sur $]0,1]$ à partir de laquelle on peut fabriquer bien d'autres lois, à commencer par la loi exponentielle : si U est uniforme et si $a>0$ on considère

$$X=-\frac{1}{a}\log U$$

qui satisfait donc si $x>0$

$$\Pr(X>x)=\Pr(U<e^{-ax})=e^{-ax}.$$

Cette loi est très utile et apparait dans les phénomènes aléatoires sans mémoire. Je donne un exemple simple, que le professeur de physique contestera un peu sans doute : on peut dire en première approximation qu'une ampoule électrique ne s'use pas. Et pourtant elle n'est pas éternelle, elle a une durée de vie aléatoire X qui est finie. Simplement, dire qu'elle ne s'use pas est dire que même si elle a duré déjà 500 heures, la probabilité qu'elle se grille dans les 10 prochaines heures est exactement la même que si elle était neuve.

Cherchons les lois possibles de X pour qu'il en soit ainsi. L'absence de mémoire se traduit ainsi : supposons qu'on sache que X est $> x$ et qu'on considère la probabilité que X soit $> x+y$ (avec x et $y > 0$). Alors cette probabilité *conditionnelle* doit satisfaire

$$\Pr(X > x + y \text{ sachant que } X > x)$$

$$= \Pr(X > x + y | X > x) = \Pr(X > y.)$$

Rappel : si A et B sont deux évènements avec $\Pr(B) > 0$ alors par définition

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

L'égalité précédente s'écrit

$$\begin{aligned}\Pr(X > x+y|X > x) &= \frac{\Pr(\{X > x+y\} \cap \{X > x\})}{\Pr(X > x)} \\ &= \frac{\Pr(X > x+y)}{\Pr(X > x)} \\ &= \Pr(X > y.)\end{aligned}$$

Introduisons $G(x) = \Pr(X > x)$. On a donc la formule magique

$$G(x+y) = G(x)G(y).$$

Si on postule que $x \mapsto G(x)$ est à valeurs dans $]0, 1[$ on en déduit vite qu'il existe un nombre $a > 0$ tel que $G(x) = e^{-ax}$, c'est bien la loi exponentielle de paramètre a.

Voici quelques détails de cette dérivation pour les amoureux de la rigueur : puisque $0 < G(1) < 1$ par hypothèse, soit $a > 0$ tel que $G(1) = e^{-a}$. Si p et q sont des entiers > 0 l'équation fonctionnelle appliquée à $x = (p-1)/q$ et $y = 1/q$ donne $G((p-1)/q)G(1/q) = G(p/q)$ ce qui par un raisonnement par récurrence sur p donne

$$G(p/q) = G(1/q)^p.$$

Faisons $p = q$: on en tire $G(1/q) = G(1)^{1/q} = e^{-a/q}$. Revenons à p quelconque, on obtient

$$G(p/q) = e^{-ap/q}.$$

Comme G est décroissante on en tire $G(x) = e^{-ax}$ comme promis.

En attendant l'autobus. Si on est assez loin de la gare routière, les instants où passe l'autobus obéissent à une loi bien connue : je prends $U_1, \ldots, U_n, \ldots$ indépendantes et de même loi uniforme sur $[0,1]$ puis $X_n = -\frac{1}{a} \log U_n$ où la constante a dépend des caractéristiques de la ligne, en fait

$$\frac{1}{a} = \mathbb{E}(X_n), \qquad \Pr(X_n > x) = e^{-ax}.$$

On forme alors

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Les

$$0 = S_0 < S_1 < S_2 \ldots < S_n < \ldots$$

sont les temps aléatoires de passage de l'autobus. Disons que l'origine 0 est celle du premier passage, très tôt le matin.

J'arrive à l'arrêt de l'autobus à l'instant $t > 0$, et je note par $K(t) \geq 1$ l'entier aléatoire défini par

$$S_{K(t)-1} < t < S_{K(t)}.$$

Ce qui signifie que je vais attendre cet autobus pendant le temps aléatoire

$$T(t) = S_{K(t)} - t < S_{K(t)} - S_{K(t)-1} = X_{K(t)}$$

La chose extraordinaire que je veux commenter est le fait que la loi de $T(t)$ est la même que celle des X_n, c'est à dire que

$$\Pr(T(t) > x) = e^{-ax}.$$

Cela est lié au fait que $X_{K(t)}$, qui majore $T(t)$, n'a plus du tout la même loi que les X_n, simplement par le fait que l'indice $n = K(t)$ est aléatoire- tout en dépendant de la suite $X_1, \ldots, X_n, \ldots$ ce qui change profondément sa loi.

On a mis longtemps à comprendre ce phénomène, mais les outils clairs que le calcul des probabilités met à notre disposition permettent de le démystifier. Pour le montrer, je vous demande d'accepter le résultat calculatoire suivant, que vos professeurs vous feraient montrer par récurrence : si $n > 0$ alors

$$\Pr(S_n < t) = \frac{a^n}{(n-1)!} \int_0^t e^{-as} s^{n-1} ds.$$

On pourrait calculer explicitement cette intégrale mais le résultat n'est utile que numériquement, et pas pour le raisonnement théorique que nous allons faire.

Prochaine étape pour $k > 1$, et $x > 0$ le calcul de

$$\begin{aligned}
&\Pr(S_{K(t)} - t > x \,; K(t) = k) \\
&= \Pr(S_{k-1} < t < t + x < S_{k-1} + X_k) \\
&\overset{*}{=} \frac{a^{k-1}}{(k-2)!} \int_0^t e^{-as} s^{k-2} \Pr(X_k > t + x - s) ds \\
&= \frac{a^{k-1}}{(k-2)!} \int_0^t e^{-as} s^{k-2} e^{-a(t+x-s)} ds \\
&= e^{-a(t+x)} \frac{a^{k-1}}{(k-2)!} \int_0^t s^{k-2} ds \\
&= e^{-a(t+x)} \frac{a^{k-1}}{(k-2)!} \left[\frac{s^{k-1}}{k-1} \right]_0^t \\
&= e^{-a(t+x)} \frac{(at)^{k-1}}{(k-1)!}.
\end{aligned}$$

Pour $k = 1$ cette égalité

$$\Pr(S_{K(t)} - t > x \,; K(t) = k) = e^{-a(t+x)} \frac{(at)^{k-1}}{(k-1)!}$$

est toujours vraie et est immédiate à démontrer.

Ceci fait tout devient très facile, on décompose l'événement $\{S_{K(t)} - t > x\}$ en tranches disjointes couvrant toutes les valeurs possibles de $K(t)$:

$$\bigcup_k \{S_{K(t)} - t > x \; ; K(t) = k\}.$$

Comme nous venons de calculer la probabilité de chaque tranche on en déduit

$$\begin{aligned}
&\mathrm{Pr}(S_{K(t)} - t > x) \\
&= \sum_k \mathrm{Pr}(S_{K(t)} - t > x \; ; K(t) = k) \\
&= \sum_k e^{-a(t+x)} \frac{(at)^{k-1}}{(k-1)!} \\
&= e^{-a(t+x)} \left(1 + at + \frac{(at)^2}{2} + \frac{(at)^3}{6} + \frac{(at)^4}{24} + \dots\right) \\
&= e^{-a(t+x)} e^{at} = e^{-ax}
\end{aligned}$$

Pour vos professeurs : le même calcul un peu perfectionné permet de montrer que si on introduit la variable aléatoire $V(t) = t - S_{K(t)-1}$ qui est le temps écoulé depuis le passage du dernier autobus, celui que j'ai raté, alors cette variable aléatoire est indépendante de $T(t)$, un résultat bien surprenant. Plus précisément, si $0 < v < t$

$$\Pr(V(t) < v \; ; \; T(t) > x) = e^{-ax}(1 - e^{-av})$$

La loi de $V(t)$ n'est ni continue, ni discrète : elle a la densité ae^{-av} sur $[0, t]$ mais

$$\Pr(V(t) = t) = e^{-at}.$$

Si t est grand, $V(t)$ est quasiment aussi de loi exponentielle. Elle est d'espérance $\frac{1}{a}(1-e^{-at})$. Et comme on a

$$X_{K(t)} = (S_{K(t)}-t)+(t-S_{K(t)-1}) = T(t)+V(t)$$

l'espérance de $X_{K(t)}$ est la somme des espérances de $T(t)$ et de $V(t)$, soit $\frac{1}{a}(2 - e^{-at})$ presque deux fois $1/a$ que nous donnait notre intuition.

Merci de votre attention.
Bonne chance pour le baccalauréat et les concours