

Semaine 22 - Probabilités et dénombrement

Valentin De Bortoli
email : valentin.debortoli@gmail.com

1 Permutations et probabilités

Après un nombre n impair de minutes la signature de la permutation de S_3 associée au changement entre le temps 0 et le temps n est impaire et donc on ne peut obtenir l'identité comme permutation au bout d'un temps impair. On suppose maintenant que l'on observe la configuration au bout d'un temps pair, $2n$. On a donc un élément de \mathcal{A}_3 (le groupe alterné de S_3). Les éléments de ce groupe sont Id , $(1\ 2\ 3)$ et $(1\ 3\ 2)$. On montre par récurrence la propriété suivante pour tout $n \in \mathbb{N}^*$: "au temps $2n$ on a une probabilité uniforme d'observer un élément de \mathcal{A}_3 ". Je détaille ici le raisonnement :

- Initialisation : entre le temps 0 et 2 on a 9 configurations différentes (3 transpositions \times 3 transpositions). Le seul moyen de revenir à l'identité est que la deuxième transposition soit l'inverse de la première soit 3 possibilités. Pour obtenir $(1\ 2\ 3)$ on peut procéder via $(12)(23)$ ou $(13)(21)$ ou $(13)(12)$. Les autres possibilités donnent $(1\ 3\ 2)$. Donc on a bien une probabilité uniforme d'obtenir un élément de \mathcal{A}_3 .
- Hérité : pour calculer la distribution de la permutation X_{2n} on utilise la relation suivante :

$$P(X_{2n} = Id) = P(X_{2(n-1)} = Id)P_{X_{2(n-1)}=Id}(X_{2n} = Id) + P(X_{2(n-1)} = (1\ 2\ 3))P_{X_{2(n-1)}=(1\ 2\ 3)}(X_{2n} = Id) + P(X_{2(n-1)} = (1\ 3\ 2))P_{X_{2(n-1)}=(1\ 3\ 2)}(X_{2n} = Id) \quad (1)$$

L'hypothèse de récurrence permet d'écrire :

$$P(X_{2n} = Id) = \frac{1}{3} \left(P_{X_{2(n-1)}=Id}(X_{2n} = Id) + P_{X_{2(n-1)}=(1\ 2\ 3)}(X_{2n} = Id) + P_{X_{2(n-1)}=(1\ 3\ 2)}(X_{2n} = Id) \right) \quad (2)$$

Les mêmes calculs que pour l'initialisation permettent de conclure que $P(X_{2n} = Id) = \frac{1}{3}(\frac{1}{3} + \frac{1}{3} + \frac{1}{3}) = \frac{1}{3}$. Il en va de même pour les autres éléments de \mathcal{A}_3 .

Ainsi on peut conclure que la probabilité est nulle si n est impair et égale à $\frac{1}{3}$ sinon.

2 Nombre de dérangements

On considère l'espace des permutations de $\llbracket 1, n \rrbracket$. On appelle dérangement toute permutation sans point fixe. Notre but ici est de donner une formule explicite du nombre de dérangements de $\llbracket 1, n \rrbracket$.

1 Dans le cas où on a deux éléments la formule devient $\text{card}(A \cup B) = \text{card}(A) + \text{card}(B) - \text{card}(A \cap B)$. La formule du crible est simplement une extension de cette formule au cas où $n > 2$.

2 On note A_i le nombre de permutations de $\llbracket 1, n \rrbracket$ telles que i est point fixe. Les dérangements constituent l'intersection des complémentaires des A_i . Donc les dérangements constituent le complémentaire de l'union des A_i . Donc on a :

$$\begin{aligned} D_n &= n! - \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 \leq \dots \leq i_k \leq n} \text{card} \left(\bigcap_{i_j \in \{i_1, \dots, i_k\}} A_{i_j} \right) \\ &= n! + \sum_{k=1}^n (-1)^k \binom{n}{k} \text{card} \left(\bigcap_{i_j \in \{1, \dots, k\}} A_{i_j} \right) \\ &= n! + \sum_{k=1}^n (-1)^k \binom{n}{k} (n-k)! \\ &= n! \sum_{k=0}^n (-1)^k \frac{1}{k!} \end{aligned} \quad (3)$$

3 Soit $D_{n,r}$ le nombre de permutations ayant exactement k points fixes. Il s'agit de choisir ces r points fixes dans $\llbracket 1, n \rrbracket$ et ensuite de compter le nombre de dérangements pour "compléter les trous". On obtient : $D_{n,r} = \binom{r}{n} D_{n-r}$

4 Il s'agit de la probabilité de tirer un dérangement en choisissant aléatoirement de manière uniforme une permutation. Lorsque le nombre de participants est grand ce nombre tend vers $\frac{1}{e}$ (résultat usuel sur les séries et l'exponentielle). Or $\frac{1}{e} \approx 0.37$. Donc on a 40 pourcents de chance de tomber sur un dérangement.

3 Suites croissantes (1)

Soit $n \in \mathbb{N}^*$ et $p \leq n$.

1 Il suffit de remarquer qu'une suite strictement croissante est caractérisée uniquement par p points parmi n . Si deux suites de p points strictement croissantes sont constituées des mêmes points alors elles sont égales. Donc le nombre de suites strictement croissantes est $\binom{p}{n}$. Il suffit de diviser par n^p pour obtenir la proportion voulue.

2 L'astuce est de poser $y_k = x_1 + \dots + x_k$. On a une suite strictement croissante de $\llbracket 1, n \rrbracket$. De plus toute suite croissante de $\llbracket 1, n \rrbracket$ (y_k) s'identifie à une suite telle que $x_1 + \dots + x_p \leq n$. En effet il suffit de poser $x_k = y_k - y_{k-1}$ si $k > 1$ et $x_1 = y_1$ sinon. Par bijection entre ces deux ensembles on a le nombre des suites qui vérifient une telle propriété qui vaut $\binom{p}{n}$

3 Il suffit de considérer les suites qui vérifient $x_1 + \dots + x_p \leq n$ mais pas $x_1 + \dots + x_p \leq n - 1$. Il faut donc retirer $\binom{p}{n-1}$ éléments de nos $\binom{p}{n}$ suites. On obtient donc $\binom{p-1}{n-1}$ suites.

4 Suites croissantes (2)

Soit $n \in \mathbb{N}^*$ et $p \leq n$.

1 Voir question précédente

2 Toute suite croissante de $\llbracket 1, n \rrbracket$, x_k , peut s'identifier à une suite strictement croissante de $\llbracket 1, n + p - 1 \rrbracket$ en posant $y_k = x_k + k - 1$. Réciproquement. Toute suite strictement croissante de $\llbracket 1, n + p - 1 \rrbracket$ peut s'identifier à une suite croissante de $\llbracket 1, n \rrbracket$. En effet on montre par récurrence que $x_k = y_k - k \in \llbracket 1, n \rrbracket$ et que $x_{k+1} \geq x_k$. Via cette bijection on a donc $\binom{p}{n+p}$ suites croissantes de $\llbracket 1, n \rrbracket$.

5 Le paradoxe du prince de Toscane

Une correction détaillée et une explication du paradoxe est disponible sur <http://www.bibmath.net/dico/index.php?action=affiche&quoi=./p/partoscane.html>.

6 Paradoxe des anniversaires

1 Une bonne idée est de regarder l'évènement contraire. On trouve que la probabilité recherchée vaut $1 - \frac{365!}{335!365^{30}}$.

2 $E(C_n) = \binom{30}{n} \frac{1}{365^{n-1}}$. En effet $C_n = \sum_{1 \leq i_1 < i_2 < \dots < i_n \leq 30} \mathbb{1}_{A_{i_1}=A_{i_2}=\dots=A_{i_n}}$ où A_k est la date d'anniversaire de l'élève

k . Les variables aléatoires étant supposées indépendantes $E(\mathbb{1}_{A_{i_1}=A_{i_2}=\dots=A_{i_n}}) = \frac{1}{365^{n-1}}$. On conclut ensuite par linéarité de l'espérance.

7 Paradoxe de Feller

1 Cette égalité correspond à un simple changement de variable lorsque l'on connaît un peu de théorie de la mesure. Ici on produit une démonstration qui ne nécessite pas cette théorie .

$$\begin{aligned} \sum_{k=0}^{n-1} P(M > k) &= \sum_{k=0}^{n-1} \left(\sum_{j=k+1}^{n-1} P(M = j) + P(M > n-1) \right) \\ &= nP(M > n-1) + \sum_{k=0}^{n-1} kP(M = k) \end{aligned} \quad (4)$$

Le terme de droite tend vers l'espérance. Si l'espérance est finie alors $\sum_{k=n}^{+\infty} kP(M = k)$ tend vers 0 lorsque n tend vers l'infini. Donc en minorant cette quantité par $n \sum_{k=n}^{+\infty} P(M = k) = nP(M > n-1)$ on trouve le résultat voulu. Si l'espérance est infini c'est encore plus simple puisqu'alors : $\sum_{k=0}^{n-1} kP(M = k)$ tend vers l'infini lorsque n tend vers l'infini. Il en va donc de même pour $\sum_{k=0}^{n-1} P(M > k)$. Ainsi dans tous les cas on a égalité.

2 Il faut donc simplement calculer $P(M > k)$. Pour se trouver dans le cas $M > k$ cela implique que parmi les $k+1$ expériences menées la première a le plus gros "score". Les expériences étant indépendantes et de même loi on trouve $P(M > k) = \frac{1}{k+1}$. En effet $P(\text{score maximum atteint en } 1) = P(\text{score maximum atteint en } 2)$ et les événements sont disjoints (c'est un point un peu subtil bien que très logique. Il faut supposer que la loi du temps d'attente est continue sinon on peut avoir des cas où le temps d'attente 1 est égal au temps d'attente 2. Pourquoi ?). Ainsi $1 = (k+1)P(\text{score maximum atteint en } 1) = (k+1)P(M > k)$. Ainsi $E(M) = \sum_{k \in \mathbb{N}^*} \frac{1}{k+1} = +\infty$.

8 Paradoxe des trois portes

1 On note A la porte que l'on a choisi. On note B la porte non-ouverte par le présentateur. On note A_W l'évènement de gain (respectivement A_L pour la perte) pour A et B_W l'évènement de gain pour B (respectivement B_L pour la perte). La même syntaxe est adoptée pour C .

$$\begin{aligned} P(B_W)_{C_L} &= \frac{P(B_W \cap C_L)}{P(C_L)} \\ &= \frac{P(B_W)}{\frac{2}{3}} \\ &= \frac{\frac{1}{3}}{\frac{2}{3}} \\ &= \frac{1}{2} \end{aligned} \quad (5)$$

Le calcul effectué ci-dessus indique que changer de porte ou ne pas en changer ne modifiera pas notre probabilité de gain. Mais il est important de remarquer que ce calcul est **faux**. En effet, on ne conditionne pas par rapport à $C_L =$ la porte C est perdante parmi les trois portes mais par rapport à $C_{L_p} =$ la porte C est perdante et est ouverte par le présentateur ce qui est très différent. En effet, $P(C_{L_p} \cap B_W) = P(B_W) = \frac{1}{3}$ ne change pas, mais $P(C_{L_p}) = P(B_L \cap C_L \cap C_{L_p}) + P(B_W \cap C_L \cap C_{L_p}) + P(B_L \cap C_W \cap C_{L_p}) = \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 1 + \frac{1}{3} \times 0 = \frac{1}{2}$. Donc on a : $P(B_W)_{C_{L_p}} = \frac{2}{3}$. Il faut donc changer de porte !

9 Paradoxe des familles

On note $G(i)$ l'évènement "l'enfant i est un garçon".

$$1 \quad P(G(1) \cap G(2))_{G(1)} = \frac{P(G(1) \cap G(2))}{P(G(1))} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

2 $P(G(1) \cap G(2))_{G(1) \cup G(2)} = \frac{P(G(1) \cap G(2))}{P(G(1) \cup G(2))} = \frac{\frac{1}{4}}{P(G(1)) + P(G(2)) - P(G(1) \cap G(2))} = \frac{1}{3}$. Une manière simple de voir cette modification (et également une manière d'envisager les probabilités conditionnelles) est de ne plus considérer des probabilités conditionnelles mais des probabilités "usuelles" pour lesquelles on aurait changé l'espace d'état. Par exemple pour la première question l'espace d'état est : $\{GG, GF\}$ et donc on a une probabilité $\frac{1}{2}$ de tirer GG . Par contre dans la deuxième question l'espace d'états est $\{GG, GF, FG\}$ et on a une chance sur trois de tirer GG . Cette approche permet de mettre de l'intuition derrière les calculs. C'est exactement le même principe pour la question suivante qui est encore plus contre-intuitive.

3 (erreur de ma part de considérer le 29 février qui rajoute une difficulté supplémentaire... Considérons simplement le 13 juin). On note $D(i)$ l'évènement "l'enfant i est né un 13 juin".

$$\begin{aligned} P(G(1) \cap G(2))_{(G(1) \cap D(1)) \cup (G(2) \cap D(2))} &= \frac{P((G(1) \cap G(2) \cap D(1)) \cup (G(1) \cap G(2) \cap D(2)))}{P((G(1) \cap D(1)) \cup (G(2) \cap D(2)))} \\ &= \frac{P((G(1) \cap G(2) \cap D(1))) + P((G(1) \cap G(2) \cap D(2))) - P(G(1) \cap G(2) \cap D(1) \cap D(2))}{P(G(1) \cap D(1)) + P(G(2) \cap D(2)) - P(G(1) \cap G(2) \cap D(1) \cap D(2))} \end{aligned} \quad (6)$$

On a : $P(G(1) \cap D(1)) = P(G(2) \cap D(2)) = \frac{1}{2 \times 365}$, $P(G(1) \cap G(2) \cap D(1) \cap D(2)) = \frac{1}{4 \times 365 \times 365}$ et $P(G(1) \cap G(2) \cap D(1)) = \frac{1}{4 \times 365}$. Donc :

$$P(G(1) \cap G(2))_{(G(1) \cap D(1)) \cup (G(2) \cap D(2))} = \frac{\frac{1}{2} - \frac{1}{4 \times 365}}{1 - \frac{1}{4 \times 365}} = \frac{729}{1459} \approx \frac{1}{2} \quad (7)$$

Encore une fois je donne aussi l'approche "changement de l'espace d'états". Pour la combinaison GG on a $364 + 365 = 729$ possibilités (en effet si le premier garçon est né un autre jour que le 13 juin cela impose que le second est né le 13 juin, donc cela représente 364 possibilités. Si le premier garçon est né le 13 juin alors le second garçon peut être né n'importe quel autre jour de l'année soit 365 possibilités). Pour la combinaison GF on a 365 possibilités tout comme pour la combinaison FG. Au final GG représente 729 possibilités parmi $2 \times 365 + 729 = 1459$ possibilités soit exactement la probabilité calculée.

Pourquoi ? On peut se demander pourquoi on obtient une quantité si compliquée pour un calcul *a priori* simple. On est tenté de dire que l'on a indépendance entre la date d'anniversaire et le sexe de l'individu mais il faut se raviser et bien écrire les choses (que ce soit en terme d'espace d'états ou avec des calculs). L'information donnée crée justement un lien entre la date de naissance et le sexe de l'individu. Par exemple étant donné que je sais qu'au moins un des deux enfants est un garçon né le 13 juin si on me donne l'information supplémentaire que les deux enfants sont de sexe opposé alors je connais la date de naissance du garçon ! Cette idée nous amène à considérer une autre question : pourquoi la probabilité est proche de $\frac{1}{2}$? Lorsque l'on sait simplement qu'un des deux enfants est un garçon la probabilité est de $\frac{1}{3}$. Lorsque l'on sait que l'un des deux enfants est un garçon **et que l'on sait lequel** (par exemple en fixant l'ainé) alors la probabilité est de $\frac{1}{2}$. Il convient de remarquer qu'en spécifiant la date d'anniversaire du garçon j'ai "quasiment" spécifié l'enfant. En effet si je sais que les deux enfants sont nés le 21 juillet et le 13 juin alors je sais que celui né le 13 juin est un garçon. Pourquoi n'obtient pas $\frac{1}{2}$ alors ? La petite différence avec le cas où l'enfant est spécifié (i.e le cas de la première question) réside dans le fait que les deux enfants peuvent être nés le même jour ! Si on refait les calculs en imposant que les enfants sont nés des jours différents alors on retombe sur $\frac{1}{2}$.

10 Un calcul de probabilité

1 $P(A)P(B) - P(A \cap B) = (x + y)(x + z) - x = yz + x(x + y + z - 1) = yz + x(1 - t - 1) = yz - xt$. Or $yz - xt \leq yz \leq y(1 - x - t - y) \leq y(1 - y) \leq \frac{1}{4}$. Même chose pour $xt - yz$. On peut conclure.

11 Indépendance et évènements quelconques

Petite erreur typographique. Il faut considérer les variables X et Y non constantes...

1 On considère X un lancer de pièce à valeurs $\{0, 1\}$ et de même pour Y . On suppose les lancers indépendants. On pose $Z = (X, Y)$. Il est facile de vérifier que X et Y sont indépendants conditionnellement à Z . Par contre Z n'est pas indépendant de X et n'est pas indépendant de Y .

2 On va étudier la quantité $A(x, y) = (P_{Z=1}(X = x) - P(X = x))(P_{Z=1}(Y = y) - P(Y = y))$. On a :

$$\begin{aligned}
A(x, y) &= P_{Z=1}(X = x, Y = y) + P(X = x, Y = y) - P_{Z=1}(X = x)P(Y = y) - P_{Z=1}(Y = y)P(X = x) \\
&= (1 - 2p(Z = 1) + p(Z = 1)^2)P_{Z=1}(X = x, Y = y) + (1 - P(Z = 1))^2 P_{Z=0}(X = x, Y = y) + \\
&\dots P(Z = 1)(1 - P(Z = 1))(P_{Z=1}(X = x)P_{Z=0}(Y = y) + P_{Z=1}(X = x)P_{Z=0}(Y = y)) - \\
&\dots P_{Z=1}(X = x)P_{Z=0}(Y = y)(1 - P(X = 1)) - P_{Z=0}(X = x)P_{Z=1}(Y = y)(1 - P(X = 1)) \\
&= (1 - P(Z = 1))^2 (P_{Z=1}(X = x) - P_{Z=0}(X = x))(P_{Z=1}(Y = y) - P_{Z=0}(Y = y))
\end{aligned} \tag{8}$$

Mais on a aussi :

$$\begin{aligned}
A(x, y) &= P_{Z=1}(X = x, Y = y) + P(X = x, Y = y) - P_{Z=1}(X = x)P(Y = y) - P_{Z=1}(Y = y)P(X = x) \\
&= (1 - P(Z = 1))P_{Z=1}(X = x, Y = y) + (1 - P(Z = 1))P_{Z=0}(X = x, Y = y) - \\
&\dots (1 - P(Z = 1))P_{Z=1}(X = x)P_{Z=0}(Y = y) - (1 - P(Z = 1))P_{Z=1}(Y = y)P_{Z=0}(X = x) \\
&= (1 - P(Z = 1))(P_{Z=1}(X = x) - P_{Z=0}(X = x))(P_{Z=1}(Y = y) - P_{Z=0}(Y = y))
\end{aligned} \tag{9}$$

Mais, puisque $P(Z = 1) \neq 1$ (sinon Z est constante et le résultat est vrai), cela impose que $P_{Z=1}(X = x) = P_{Z=0}(X = x)$ ou $P_{Z=1}(Y = y) = P_{Z=0}(Y = y)$ et donc $P_{Z=1}(X = x) = P(X = x)$ ou $P_{Z=1}(Y = y) = P(Y = y)$. Supposons qu'il existe x tel que $P_{Z=1}(X = x) \neq P_{Z=0}(X = x)$ alors on a pour tout y l'égalité $P_{Z=1}(Y = y) = P_{Z=0}(Y = y)$. Donc Z indépendant de Y . De même, si il existe y tel que $P_{Z=1}(Y = y) \neq P_{Z=0}(Y = y)$. On peut donc conclure dans le cas où Z est binaire.

Remarque : cet exercice est assez difficile, voici quelques conseils :

- montrer une relation d'indépendance revient à vérifier une égalité algébrique. Il faut absolument réussir à coder toute l'information de l'énoncé dans une équation.
- pour le premier calcul on a raisonné en découpant $P(X = x, Y = y) = P(X = x)P(Y = y)$ et en examinant les différents cas pour X et pour Y .
- pour le second calcul on a raisonné en ne découpant pas $P(X = x, Y = y)$ mais en considérant les différents cas pour le couple (X, Y)