M2 MEEF. Approfondissement disciplinaire : Probabilités et combinatoire

Cette séance se déroulera en trois parties :
- une première partie sur la modélisation en probabilités, et plus particulièrement sur les probabilités conditionnelles.
- une seconde partie sur des modèles de tirages et les problèmes de dénombrement associés.
- pour finir, un certains nombre de problèmes divers, pouvant faire appel à une modélisation mathématique ou une résolution numérique, sont proposés. Vous pourrez choisir entre une approche numérique (c'est-à-dire une simulation des situation proposées) ou une approche combinatoire (menant à un calcul exact).

Vous travaillez en groupe de 2 ou 3 étudiants. Vos travaux sur les questions de la partie "Modélisation : les probabilités conditionneles" seront à rendre sur feuille (une par groupe) et seront notés.

La majorité des problèmes divers présentés ici provient du livre [3]. Nous recommandons, sur le thème des paradoxes en probabilité, l'ouvrage de L. Schneps et C. Colmez [1].

1 Modélisation : les probabilités conditionnelles

Les lois conditionnelles apparaissent quand on dispose d'une variable aléatoire, et que l'on opère une sélection sur cette variable. Nous allons en donner deux exemples.

1.1 Le dé et l'interprétation fréquentiste

On lance un dé à à 6 faces équilibré. On note X le résultat de ce lancer.

1. Décrivez l'univers et la loi de probabilité $\mathbb{P}$ sur cet univers.
2. Rappelez l'interprétation fréquentiste de l'énoncé $\mathbb{P}(X = 6) = 1/6$.
3. Calculez $\mathbb{P}(X = 6 \mid X \text{ est pair})$, puis énoncez-en une interprétation fréquentiste.
4. Démontrez l'interprétation fréquentiste précédente.

1.2 Le bombardier et le biais du survivant

Parfois, on dipose a priori de plusieurs modèles pour une même situation. Le modèle choisi peut grandement affecter les conclusions que l'on peut en tirer. Ceci est le cas, par exemple, avec le biais du survivant, dont nous allons donner un exemple historique simplifié.

Lors de la deuxième guerre mondiale, l'armée de l'air des États-Unis mène des raids sur le territoire allemand. Au cours de ces raids, certains avions sont touchés par la DCA, causant parfois la perte de l'appareil. Afin de diminuer les pertes, il est proposé de renforcer les avions. Cependant, comme de tels renforts alourdissent l'avion et réduisent ses capacités, on ne peut renforcer que les endroits les plus critiques. On mène une analyse statistique, et dans un premier temps, on compte les points des avions les plus touchés par les impacts de DCA.

Une première approche naïve consiste à adopter un modèle probabiliste simple : les avions sont touchés par la DCA, et on observe directement ces impacts. Afin de mieux protéger les appareils, il faut alors renforcer les zones les plus touchées, c'est-à-dire présentant le plus d'impacts.

Une deuxième approche, proche de celle adopté par A. Wald du *Stastical Research Group* [2], est la suivante :
- On suppose que les impacts de DCA sont répartis uniforméments sur les avions.
- Ensuite, suivant la localisation de l'impact, l'avion peut, ou non, se crasher.
- On n'observe les impacts que sur les avions revenus de mission ; autrement dit, la loi des impacts est la loi uniforme *conditionnée par le fait que l'avion a pu revenir se poser*.

Si l'on suit les conclusions de ce modèle, il vaut mieux renforcer les parties de l'avion sur lesquelles on observe peu ou pas d'impacts !

Cet exemple est, historiquement, l'un des premiers pour lesquels le *biais du survivant* a été mis en évidence. Il est très important de prendre ce biais en compte dans certaines analyses statistiques. Par exemple, si l'on veut évaluer l'effet d'un certain traitement médical après 10 ans, il faut prendre en compte que les personnes observées après 10 ans sont celles qui ne sont pas mortes entre temps. On observe donc la santé des patients *conditionnée par le fait d'avoir survécu au moins* 10 *ans*.

1.3 Les boîtes de Bertrand

Maintenant, à vous ! Le problème des boîtes de Bertrand est un classique d'application des probabilités conditionnelles. Dans le problème suivant :
- modélisez la situation ;
- dessinez un arbre des probabilités ;

— répondez à la question qui vous est posée.

Trois boîtes opaques sont posées devant vous. On vous dit que l'un des boîtes contient 1000 pièces de cuivre, une autre 999 pièces de cuivre et 1 pièce d'argent, et la troisième 1000 pièces d'argent.

Vous choisissez une boîte au hasard, et vous piochez une pièce au hasard dans cette boîte. Vous obtenez une pièce d'argent.

Si vous tirez une deuxième pièce au hasard (sans remise) dans la même boîte, de quel métal sera-t-elle faite ?

1.4 Garçon ou fille ?

Le but du problème suivant est d'analyser un soi-disant paradoxe, parfois présenté comme une application non intuitive des probabilités. Comme nous allons le voir, la situation est plus complexe, et dépend de façon subtile de la modélisation adoptée (et en particulier des évènements par lesquels on conditionne).

Pour chacun des deux problèmes ci-dessous :
- modélisez la situation ;
- dessinez un arbre des probabilités ;
- répondez à la question qui vous est posée.

Vous prendrez grand soin de bien traduire dans votre modélisation l'évènement utilisé pour conditionner.

1.4.1 Première variante

Vous croisez au marché un collègue accompagné d'un garçon, qui est son fils. Vous savez que ce collègue a deux enfants. Vous voudriez prendre des nouvelles de l'autre enfant, mais ne vous rappelez plus si c'est un garçon ou une fille.

Entre "Comment va ta fille ?" et "Comment va ton autre garçon ?", y a-t-il une question qui vous donne plus de chances de tomber juste que l'autre ?

1.4.2 Deuxième variante

Le roi de France, qui vient de mourir, a eu deux enfants. Le Dauphin, désigné selon la loi salique, sera bientôt couronné.

À votre avis, le Dauphin a-t-il un frère ou une sœur ? Une des deux réponses est-elle plus probable que l'autre ?

1.5 Un peu de génétique

Une des nombreuses applications des probabilités conditionnelles concerne le calcul du risque de maladies génétiques. En voici un exemple simple.

Un gène codant pour des récepteurs rétiniens a deux allèles : un allèle D dominant et un allèle d récessif. La fréquence de l'allèle d récessif est de 10%.

Ce gène est porté par le chromosome sexuel X. Une femme a deux copies de ce gène. Si elle est porteuse de la combinaison DD ou Dd, sa vision est normale ; si elle est porteuse de la combinaison dd, elle est daltonienne.

1. Le génotype d'une femme est DD, Dd ou dd. Quelle est la loi de ces différentes combinaisons ?

Un homme a une seule copie de ce gène. S'il porte l'allèle D, sa vision est normale ; s'il porte l'allèle d, il est daltonien.

Marie a deux garçons, Alfred et Balthazar. Elle a transmis à chacun de ses enfants une copie de l'un de ses deux chromosomes X. Alfred et Balthazar sont donc chacun porteurs d'un allèle D ou d. On s'intéresse au génotype de Marie, Alfred et Balthazar.

2. Modélisez le problème.
3. Alfred vous dit qu'il est daltonien. À votre avis, sa mère Marie l'est-elle ? Et son frère Balthazar ?
4. Un élève vous dit que "C'est facile, si je connais les allèles de Marie, alors ceux d'Alfred et Balthazar sont indépendants". Comment formaliseriez-vous cette affirmation ?
5. Utilisez la propriété de la question précédente pour retrouver la probabilité que Balthazar soit daltonien sachant qu'Alfred l'est.

2 Combinatoire : modèles de tirages

2.1 Tirages et modélisation

De nombreux problèmes de combinatoire élémentaires peuvent se ramener à des calculs classiques : dénombrement du nombre de parties de cardinal fixé d'un ensemble, du nombre d'applications injectives d'un ensemble dans un autre, du nombre d'applications surjectives d'un ensemble dans un autre...

Cependant, traduire un problème concret en termes d'applications entre ensembles vérifiant certaines propriétés (injectivité ou surjectivité, par exemple) est délicat, et demande de l'expérience. Afin de faciliter le raisonnement, on introduit souvent des modèles stéréotypés, typiquement des modèles de tirage. Dans une situation raisonnablement simple, le comptage du nombre de tirages possibles est équivalent à un des comptages abstraits évoqués précédemment. De plus, l'aspect concret de tels tirages permet de s'y ramener plus intuitivement.

Il existe de nombreux modèles de tirage, correspondant à des situations diverses. Dans cette partie, on se donne un ensemble fini E, de cardinal n. On peut y penser comme à des boules étiquetées dans une urne. On va tirer r boules (ou éléments de E) :

- avec ou sans remise ;
- de façon ordonnée ou non.

Un tirage est *avec remise* si l'on peut tirer plusieurs fois le même élément (on remet la boule tirée dans l'urne), et *sans remise* si l'on ne peut tirer chaque élément qu'une fois au plus. Un tirage est *ordonné* si l'on prend en compte l'ordre dans lequel on a tiré les éléments, et *non ordonné* si seul importe le nombre de fois où l'on a tiré chaque élément.

Remarque 1.
Les tirages non ordonnés sont parfois appelés simultanés : toutes les boules sont tirées en même temps. Cependant, s'il est facile de faire un tirage simultané sans remise, la notion de tirage simultané avec remise est plus gênante : on ne peut pas replacer les objets entre deux tirages si ces tirages sont simultanés. Le vocabulaire de "tirage ordonné" et "tirage non ordonné" évite cet écueil.

Nous avons donc quatre types de tirages possible : ordonné avec remise (O, R), ordonné sans remise (O, SR), non ordonné avec remise (NO, R), non ordonné sans remise (NO, SR). Commençons par analyser quelques exemples.

1. Rattachez, en le justifiant brièvement, chacun des problèmes ci-dessous à un de ces quatre types de tirage. Ne les résolvez pas maintenant !
   (a) Une main de poker est contituée de cinq cartes d'un jeu de 32 cartes. Combien y a-t-il de mains de poker possibles ?
   (b) Combien le mot "POKER" a-t-il d'anagrammes (qui ne sont pas nécessairement des mots du dictionnaires) ?
   (c) Patrick décide d'acheter 3 bouteilles de bière ; le pub en propose 45 types différents. Combien a-t-il d'assortiments possibles ?
   (d) L'une de vos classes a 28 élèves. Vous faites la liste des anniversaires des élèves. Combien de listes possibles y a-t-il ? Vous pourrez supposer qu'aucun élève n'est né une année bissextile.
   (e) Votre autre classe a 26 élèves. L'un de vos collègues vous affirme : "C'est remarquable, dans cette classe, tous les élèves ont des anniversaires différents." Pour vous en assurer, vous faites la liste des anniversaires des élèves. En supposant que votre collègue a raison, combien de listes possibles y a-t-il ?
   (f) Le lapin de Pâques dispose de 7 œufs en chocolat, colorés aux couleurs de l'arc-en-ciel. Il veut les distribuer à 3 enfants, Alice, Bob et Oscar. Combien de façons différentes a-t-il de distribuer ses friandises ?
   (g) Le Père Noël veut distribuer 20 truffes au chocolat (identiques) à Alice, Bob et Oscar. Combien de façons différentes a-t-il de distribuer ses friandises ?
   (h) Une combinaison de Loto est composée de 6 nombres distincts, choisis entre 1 et 49. Combien y a-t-il de combinaisons possibles ?

2.2 Dénombrement

Maintenant, nous allons dénombrer les possibilités pour chaque tirage.

2. Pour chacun des quatre types de tirages possibles, donner un univers décrivant l'ensemble des issues. On notera $\Omega_{O,R}$, $\Omega_{O,SR}$, $\Omega_{NO,R}$ et $\Omega_{NO,SR}$ ces univers.

Remarque 2.
Afin de travailler avec des tirages non-ordonnés avec remise, on utilise parfois des multi-ensembles. Un multi-ensemble est un ensemble (donc non ordonné), mais dont les éléments peuvent avoir une multiplicité, comme par exemple :

$$\{\{A, B, A, A\}\} = \{\{A, A, A, B\}\} \neq \{\{A, B\}\}.$$

Nous utiliserons cette notion pour le reste de cette partie ; cependant, elle est absolument hors programme !

3. Exprimez les cardinaux des univers $\Omega_{O,R}$, $\Omega_{NO,R}$ et $\Omega_{NO,SR}$ en fonction du nombre d'éléments n et du nombre de tirages r.

Il est naturel de se demander quel est le nombre de tirages non ordonnés avec remise. Pour simplifier, nous prendrons $E = \{1, \dots, n\}$. Étant donné un tirage non ordonné avec remise :

- on classe les nombres tirés par ordre croissant ;
- on introduit des "cloisons" (barres verticales $|$) entre les nombres : entre deux nombres successifs a et b, on place $b - a$ cloisons.
- on remplace les nombres tirés par des boules o.

4. On prend $E = \{1, 2, 3, 4, 5\}$ et $r = 7$. Encodez les tirages $\{\{1, 4, 2, 2, 1, 3, 5\}\}$ et $\{\{3, 4, 2, 4, 2, 5, 5\}\}$. Décodez les tirages $oo|oo|o|o|o$ et $||oo|oo|ooo$.
5. Pour $E = \{1, 2, 3, 4, 5\}$ et $r = 7$, combien y a-t-il de tirages non ordonnés avec remise ? Et pour n et r quelconques ?
6. Reprenez les exemples de la première question, et répondez à chaque question.

2.3 Oublier l'ordre

Si l'on se donne un tirage ordonné, on peut toujours oublier l'ordre pour obtenir un tirage non ordonné. Par exemple, quand on distribue une main à la belote, les cartes sont distribuées une à une ; le tirage est, à ce niveau, ordonné et sans remise. Cependant, l'ordre dans lequel les cartes ont été distribuées n'a aucune importance pour la suite du jeu. On considère donc que deux mains sont identiques si elles elles sont composées des mêmes cartes, quelque soit l'ordre de distribution : on oublie cet ordre. On obtient alors un tirage non ordonné et sans remise.

Mathématiquement, cela se traduit par l'existence d'une application $f_{SR} : \Omega_{O,SR} \to \Omega_{NO,SR}$, qui à une suite $(x_1, \dots, x_r)$ associe l'ensemble $\{x_1, \dots, x_r\}$, avec $1 \leq r \leq n$.

7. Montrez que f_{SR} est bien définie.
8. On se donne un tirage non ordonné $\omega = \{x_1, \dots, x_r\}$. Quel est le cardinal de $f_{SR}^{-1}(\omega)$?
9. On munit $\Omega_{O,SR}$ de la loi uniforme. Montrez que la loi de f_{SR} est uniforme sur $\Omega_{NO,SR}$.

Autrement dit, **si l'on fait un tirage uniforme ordonné et sans remise, puis que l'on oublie l'ordre, on obtient un tirage uniforme non ordonné et sans remise.** C'est heureux pour les joueurs de cartes : en distribuant aléatoirement et uniformément les cartes une à une, on obtient bien des mains aléatoires uniformes !

On peut faire de même avec des tirages avec remise : il existe une application $f_R : \Omega_{O,R} \to \Omega_{NO,R}$, qui à une suite $(x_1, \dots, x_r)$ associe le multi-ensemble $\{\{x_1, \dots, x_r\}\}$.

10. On prend $E = \{0, 1\}$ et $n \geq 2$. On munit $\Omega_{O,R}$ de la loi uniforme. La loi de f_R est-elle uniforme sur $\Omega_{NO,R}$?

Cette subtilité est importante en termes de modélisations : si l'on veut obtenir un tirage non ordonné avec remise uniforme, on ne peut pas simplement partir d'un tirage ordonné avec remise uniforme et oublier l'ordre du tirage !

3 Le "paradoxe" des anniversaires

Dans votre classe de trente élèves, deux ont un anniversaire qui tombe le même jour. Est-ce vraiment étonnant ? Pour répondre à cette question, on va s'intéresser plus généralement à la probabilité p_n que, dans une classe de n personnes, toutes aient des jours anniversaire différents.

Modélisez le problème. Vous aurez intérêt à faire des hypothèses simplificatrices, mais il faudra les expliciter.

ESTIMATION NUMÉRIQUE

Écrivez un programme qui effectue l'expérience aléatoire correspondante à votre modèle.

Utilisez ce programme pour estimer la valeur de p_{30}. Quelle est l'ordre de grandeur de l'erreur entre la valeur estimée et la valeur exacte de p_{30} ? Commentez la question initiale sur votre classe de trente élèves.

Utilisez votre programme pour tracer la valeur de p_n en fonction de n, pour $1 \leq n \leq 50$.

CALCUL EXACT

On va calculer (de deux manières) la valeur de p_n.

1. Première méthode : exprimez p_{n+1} en fonction de p_n. Déduisez-en une expression pour p_n.
2. Deuxième méthode : commencez par expliciter votre modèle, autrement dit décrivez l'espace $(\Omega, \mathbb{P})$. Vous allez a priori donner une probabilité $\mathbb{P}$ uniforme. Combien y a-t-il de configurations au total ? Combien de configurations correspondent à une situation où tous les élèves ont des jours anniversaire différents ? En déduire une expression pour p_n.

Calculez la valeur de $1 - p_{30}$. Commentez la question initiale sur votre classe de trente élèves. À partir de quelle valeur de n a-t-on $1 - p_n \geq 1/2$?

4 La ruine du joueur

Un homme se vante d'être allé au casino chaque été pendant dix ans, et d'y avoir à chaque fois (globalement) gagné. Il se vante donc d'avoir une chance exceptionnelle. En l'interrogeant sur sa technique, on apprend qu'à chaque visite, il arrête de jouer dès qu'il a un bénéfice de 100€, et qu'il a une mise initiale de 1000€. On va voir si une telle réussite est vraiment exceptionnelle.

On suppose donc que cet homme joue à chaque fois 10€, à un jeu où la probabilité de gagner sa mise est de $p < 1/2$, celle de perdre sa mise de $1 - p$, qu'il commence avec des fonds de 1000€, et qu'il s'arrête dès que ses fonds arrivent à zéro ou qu'il atteint la somme de 1100€qu'il s'était fixée. Pour les applications, vous prendrez $p = 0,48$ (qui est un réglage typique de machine à sous)

ESTIMATION NUMÉRIQUE

Écrivez un programme qui réalise l'expérience aléatoire en question, et qui renvoie "Gagné" si le joueur a atteint sa cible de 1100€, et "Perdu" sinon.

Estimez numériquement la probabilité que le joueur atteigne sa cible une année donnée. Et dix ans de suite ? Le résultat vous semble-t-il extraordinaire ?

CALCUL EXACT

On généralise le problème : on suppose que le joueur dispose au départ de $10z$€, avec $0 \leq z \leq 110$. On note $f(z)$ la probabilité que le joueur termine gagnant (donc avec un montant de 1100€). Que valent $f(0)$ et $f(110)$? En conditionnant par le résultat de la première partie, montrer que :

$$f(z) = pf(z+1) + (1-p)f(z-1) \qquad \forall 1 \leq z \leq a-1.$$

En écrivant $f(z)$ comme $pf(z) + (1-p)f(z)$, en déduire une équation vérifiée par $f(z) - f(z-1)$, et obtenir l'expression de $f(z)$.

Calculer la probabilité que le joueur atteigne sa cible une année donnée. Et dix ans de suite ? Le résultat vous semble-t-il extraordinaire ?

5 Loi hypergéométrique et estimation de population

Supposons que l'on veuille estimer la population de poissons d'un lac. On imagine la méthode suivante : on pêche 1000 poissons sur toute la surface du lac, et on les marque tous avant de les relâcher. On attend alors quelques jours, puis on pêche à nouveau 1000 poissons. On observe que 100 d'entre eux sont marqués. Quelle information cela nous donne-t-il sur la population totale ?

ESTIMATION NUMÉRIQUE

Écrivez un programme qui simule la variable aléatoire correspondant au nombre de poissons marqués repéchés, la population totale de poissons N étant fixée.

Utilisez ce programme pour estimer la probabilité que le nombre de poissons marqués repéchés soit entre 95 et 105 (inclus), pour des populations totales de poissons de votre choix. Pour quelle valeur (approximative) de N cette probabilité est-elle maximale ?

CALCUL EXACT

On revient encore une fois aux boules ! supposons que l'on ait n_1 boules rouges et n_2 boules noires. On en tire r au hasard. La probabilité que k d'entre elles soient rouges est $\binom{n_1}{k}\binom{n_2}{r-k}/\binom{n_1+n_2}{r}$. Justifiez ce résultat. La loi ainsi définie est appelée *loi hypergéométrique*.

Exprimez la probabilité ci-dessus que l'on ait obtenu 100 poissons sur les 1000 de la deuxième pêche. Cette probabilité dépend de la population totale $N = n_1 + n_2$. Chercher la valeur de N qui maximise cette probabilité.

On utilise cette valeur comme estimateur de N. Cette technique, qui semble parfaitement heuristique, est une méthode efficace et universelle dont on sait prouver qu'elle fonctionne dans une grande variété de situations : on calcule la probabilité qu'avait a priori l'observation que l'on a faite, en fonction du paramètre inconnu, et on estime ce paramètre par la valeur qui maximise cette probabilité. L'estimateur s'appelle l'*estimateur du maximum de vraisemblance* (puisque c'est celui qui maximise la probabilité a priori de l'évènement observé, et que cette probabilité est appelée *vraisemblance*).

6 Les clients se placent-ils au hasard au restaurant ?

On arrive dans un restaurant qui est constitué d'un long comptoir de seize places. On observe qu'il y a cinq clients, et qu'ils sont placés de telle manière qu'il y a toujours au moins une place libre entre deux places occupées. On se demande si les clients ont choisi leur place au hasard. Pour cela, il y a un peu de travail.

Supposons maintenant que r clients arrivent dans le restaurant de $n \geq r$ places, et qu'ils choisissent leur place au hasard au fur et à mesure de leur arrivée. On appelle *configuration finale* la liste des groupes de places consécutives vides ou occupées ; un exemple de configuration finale est

V, O, V, V, O, V, V, V, O, V, V, V, O, O, V, V,

où l'on a noté V les places vides et O les places occupées.

ESTIMATION NUMÉRIQUE

Écrivez un programme qui réalise l'expérience aléatoire en question avec 16 places et 5 clients, et qui renvoie le nombre de groupes de clients. Estimer numériquement la loi du nombre de groupes de clients.

Pensez-vous, en fin de compte, que les clients du restaurant dont on parlait au début de l'énoncé se sont placés au hasard ? Et si onze clients étaient répartis en trois "groupes" ?

CALCUL EXACT

Montrez que toutes les configurations finales vérifiant que le nombre total de places occupées est r sont équiprobables.

Si l'on a k "groupes" de clients (au sens où ils sont finalement assis en groupe, pas au sens où ils sont arrivés ensemble), à quoi peut ressembler la configuration finale ? En déduire la probabilité, si les clients ont choisi au hasard, que l'on ait k groupes de clients. La réponse finale est :

$$\frac{\binom{r-1}{k-1}\binom{n-r+1}{k}}{\binom{n}{r}}.$$

Calculer, lorsque $n = 16$ et $r = 5$, la probabilité que $k \geq 5$. Pensez-vous, en fin de compte, que les clients du restaurant dont on parlait au début de l'énoncé se sont placés au hasard ? Et inversement, si onze clients étaient répartis en trois "groupes" ?

Dans les deux cas, la logique est celle des tests d'hypothèses : lorsque l'on veut décider si les valeurs prises par la variable aléatoire observée sont anormales, on ne considère jamais la probabilité que la variable prenne la valeur précise (par exemple $k = 3$) mais celle que la variable prenne cette valeur "ou quelque chose de plus étonnant encore" (par exemple $k \leq 3$) [1].

7 Élections et marches aléatoires

On observe le dépouillement d'une élection dans une bureau de vote de 1000 votants. On suppose pour faire simple qu'il n'y a que deux candidats, pas d'abstention ni de bulletins blancs ou nuls. En fin de compte, le vainqueur gagne avec 560 voix contre 440, et on observe que ce vainqueur est resté en tête tout au long du dépouillement. Vos élèves vous disent que cela prouve bien que l'élection est truquée.

Pour discuter ce point, on va considérer une marche aléatoire sur $\mathbb{Z}$, qui fait des pas de $+1$ ou -1, avec des probabilités de p et $(1-p)$ respectivement, et qui part de l'origine. On va donc représenter tout résultat possible de l'expérience par un chemin du type illustré ci-dessous :

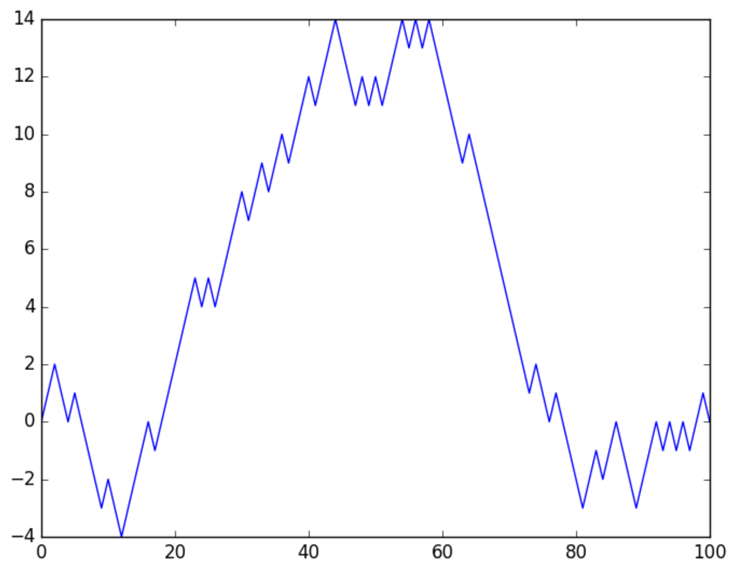

FIGURE 1 – Un exemple de marche aléatoire à 100 pas, avec $p = 1/2$.

1. Pour $n \in \mathbb{N}$ et $x \in \mathbb{Z}$, on note $N_{n,x}$ le nombre de chemins qui vont de l'origine $(0, 0)$ à (n, x) (donc en n pas). Montrer que ce nombre est nul si $n + x$ est impair, et vaut

$$N_{n,x} = \binom{n}{\frac{n+x}{2}}$$

si $n + x$ est pair.

1. à ce titre l'exemple de l'exercice est mal choisi ! on y considère en pratique la probabilité que $k = 5$... mais uniquement parce que c'est la même chose que $k \geq 5$.

2. On se donne deux nombres de pas $a < b$ et deux entiers α, β strictmeent positifs. Soient $A := (a, \alpha)$ et $B := (b, \beta)$. Soit $B' = (b, -\beta)$ le point image de B par la réflexion par rapport à l'axe des abscisses. Montrez que le nombre de chemins de A à B qui touchent (ou traversent) l'axe des abscisses est égal au nombre de chemins de A à B'. Vous pourrez utiliser l'application qui prend un chemin de A à B touchant l'axe des abscisses, et lui associe le chemin obtenu en le réfléchissant par rapport à l'axe des abscisses à partir de la première intersection.
3. Soient n et x comme ci-dessus. Montrer que le nombre de chemins de l'origine à (n, x) qui restent toujours strictement au-dessus de l'axe des abscisses est égal à
$$N_{n-1,x-1} - N_{n-1,x+1} = \frac{x}{n} N_{n,x}.$$
4. En déduire que, dans une élection entre deux candidats qui obtiennent en fin de compte v_1 et v_2 votes, avec $v_1 > v_2$, la probabilité que le premier candidat reste en tête pendant toute l'élection est $\frac{v_1 - v_2}{v_1 + v_2}$. À supposer que l'élection n'est pas truquée, quelle est cette probabilité dans le cas de l'élection discutée en introduction ? L'évènement vous semble-t-il exceptionnel ?

8 Les anniversaires, suite

La probabilité p_n que l'on a calculée dans le paragraphe sur les anniversaires, au fond, est la probabilité que sur n personnes, on ait n jours anniversaires différents. On peut chercher à en savoir un peu plus, et à calculer la probabilité $p_{n,k}$ que sur n personnes, on ait k jours anniversaires différents (on aura donc avec ces notations $p_n = p_{n,n}$).

1. Démontrez la formule d'inclusion-exclusion : si on se donne des évènements $A_1, \ldots, A_k$, alors :
$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} (-1)^{k+1} \sum_{\substack{I \subset [[1,k]] \\ |I| = k}} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)$$
On pourra utiliser le fait que $\mathbf{1}_{\bigcup_{i=1}^k A_i} = 1 - \mathbf{1}_{\bigcap_{i=1}^k A_i^c} = 1 - \prod_{i=1}^k \mathbf{1}_{A_i^c}$.
2. Combien y a-t-il de façons de placer n boules distingables dans k urnes distingables, de telle sorte que chaque urne contienne au moins une boule ? Indication : considérer les évènements $A_i = \{\text{l'urne } i \text{ est vide}\}$.
3. En déduire la loi du nombre d'anniversaires dans une classe de n élèves, c'est-à-dire calculer la probabilité d'avoir k jours anniversaires dans une classe de n élèves.

Références

[1] C. Colmez et L. Schneps, *Les Maths au tribunal. Quand les erreurs de calcul font les erreurs judiciaires*, Éditions Science Ouverte, 2015.

[2] M. Mangel et F.J. Samaniego, *Abraham's Wald work on aircraft survivability*, Journal of the American Statistical Association, Volume **79**, no. 386 (Jui. 1984), 259–267.

[3] W. Feller, *An introduction to probability theory and its applications. Vol. I*, Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.