M2 MEEF : approfondissement
Combinatoire et applications

Dans cet énoncé, on n'utilisera que des outils de combinatoire (nombre de combinaisons, permutations, *etc.*) On appliquera ces outils à des questions de prise de décision sur différents modèles. La quasi-intégralité des modèles traités ici provient du livre [1].

Les différentes sections sont rangées par ordre de difficulté croissante.

1 Garçon ou fille ?

PREMIÈRE 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 ?

DEUXIÈME 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 ?

2 Rappels de combinatoire

On rappelle quelque formules de base de la combinatoire : si l'on a n boules distinguables, et que l'on tire r boules, alors

- si l'on tire avec remise, le nombre de tirages (ordonnés) possibles est n^r ;
- si l'on tire sans remise, le nombre de tirages (ordonnés) possibles est $\frac{n!}{(n-r)!}$;
- si l'on tire sans remise, le nombre de tirages (non ordonnés) possibles est $\frac{n!}{r!(n-r)!}$.

On notera $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. On peut se poser (au moins) deux questions :

1. Sauriez-vous justifier les deux dernières expressions en partant du fait que $n!$ est le nombre de permutations de n boules distinguables ?
2. Au vu de la liste ci-dessus, il semble naturel de se demander quel est le nombre de tirages non ordonnés avec remise. Observez que décrire un tel tirage revient

à dire, dans une liste de $n + r - 1$ objets, lesquels sont des boules et lesquels sont des cloisons. Justifiez proprement la bijection sous-jacente (par exemple en associant à tout tirage un choix boules/cloisons, et à tout choix boules/cloisons un tirage). Déduisez-en le nombre recherché.

3 Le “paradoxe” des anniversaires

On commence par un classique : dans votre classe de trente élèves, deux ont un anniversaire qui tombe le même jour. Est-ce vraiment étonnant ? Pour répondre à cette quesiton, 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 programe 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 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 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 à chaque fois on les marque 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

Modéliser la situation, et écrivez un programme qui calcule 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 $\binom{n_1}{k}\binom{n_2}{r-k}/\binom{n}{r}$. Justifiez scrupuleusement ce résultat. La loi ainsi définie est appelée *loi hypergéométrique*.

Exprimer 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. Chercher le 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*).

5 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

Écrire 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.

Estimer numériquement la probabilité que le joueur atteigne sa cible une année donnée. Et dix ans de suite ? Est-ce que le résultat vous semble extraordinaire ?

CALCUL EXACT

On généralise le problème : on suppose que le joueur dispose au départ de z€, avec $0 \leq z \leq 1100$. On note $f(z)$ la probabilité que le joueur termine gagnant (donc avec un montant de 1100€). Que valent $f(0)$ et $f(1100)$? 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 ? Est-ce que le résultat vous semble extraordinaire ?

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, et 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

Écrire 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 inversement, 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. Le vainqueur gagne en fin de compte 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 :

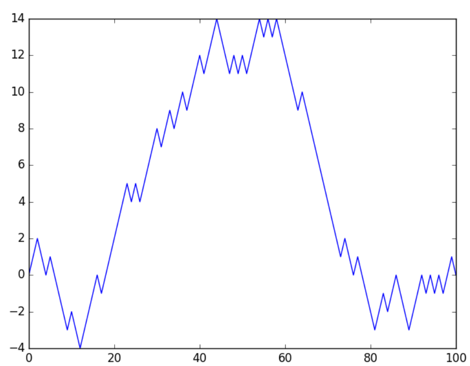

FIGURE 1 – Un exemple de marche aléatoire à 100 pas, avec $p = 1/2$.

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.
2. Soient $A = (a,\alpha)$ et $B = (b,\beta)$ deux points à coordonnées entières, avec $a < b$ et $\alpha > 0$, $\beta > 0$ et $B' = (b,-\beta)$ le point image de B par rapport à l’axe des abscisses. Montrer que le nombre de chemins de A à B qui touchent (ou traversent) l’axe des abscisses chemin de A à B qui touche l’axe des abscisses, le chemin obtenu en le retournant par rapport à l’axe à partir de la première intersection).

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 $k \geq 5$.

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 $\frac{x}{n} N_{n,x}$.
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, c'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. Démontrer la formule d'inclusion-exclusion : si on se donne des évènements A_1, ..., 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] W., Feller, *An introduction to probability theory and its applications. Vol. I,*