Probabilités

Marc SAGE

18 juin 2014

Table des matières

1 Avertissement probabilistique (paradoxe de Bertrand, version soft)

On considère deux bancs chacun de deux places. Un promeneur fatigué vient s'asseoir au hasard sur une place, puis un autre promeneur en fait de même (sans prendre la même place que le premier). *Quelles chances ont les deux promeneurs de s'asseoir sur un même banc ?* Donner au moins deux réponses distinctes (et justifiées !) puis disserter longuement sur les dangers du langage à créer des fictions qui sont par essence *hors* discours mathématique.

Solution proposée.

Les deux promeneurs étant *distinguables* (l'énoncé précise "un *autre*") et leurs assises *successives*, il est naturel de choisir comme univers un ensemble de couples. Mais nous pouvons lister *les bancs* tout aussi bien que *les places*. Dans le premier cas, une fois le premier promeneur assis, le second aura une chance sur deux de choisir le même banc. Dans le second cas, il reste trois places dont une seule sur le même banc que le premier promeneur, d'où une chance sur trois.

La question "*Quelles chances...*" se reformule proprement en "*Quelle est la probabilité mathématique que...*". Et une probabilité présuppose (du moins en mathématique) :

1. un univers modélisant notre expérience (dans le premier cas, par exemple le carré $\{A, B\}^2$ où chaque lettre code un banc ; dans le second, par exemple l'ensemble des paires injectives à valeurs dans $\{a, b, c, d\}$ où chaque lettre code une place) ;
2. une mesure de probabilité sur cette univers (on a ici arbitrairement choisi celle uniforme).

La confusion sévit ici pernicieusement à deux niveaux : d'une part, le terme *probabilité* désigne à la fois "le fait d'être probable[1]" et la quantification mathématique (numérique) de ce fait, d'autre part cette quantification est un *acte* (qui dépend de notre libre arbitre : choix double d'un univers et d'une mesure de probabilité) et non un impératif divin assurant son univocité. Ainsi l'énoncé parle-t-il d'une fiction à cheval entre deux mondes et qui n'a de consistance dans aucun d'eux : dans le monde usuel, elle n'est pas numérique, dans le monde mathématique, "sa" valeur dépend d'actes humains et non de la sacro-sainte vérité mathématique[2].

2 Paradoxe de Bertrand (version disque)

On considère un cercle de rayon 1 que l'on appelle C. On inscrit dans C un triangle équilatéral que l'on appelle T. On appelle L la longueur du côté de T. On appelle O le centre de C et D le disque associé à C. On admet que le rayon du cercle inscrit de T vaut $\frac{1}{2}$.

On s'intéresse à la probabilité (au sens commun) qu'une corde de C "choisie au hasard" ait une longueur supérieure ou égale à L.

1.
   (a) *On choisit de coder une corde par son milieu. Quel univers est-il alors pertinent de choisir ?*
   (b) *On définit la probabilité d'une partie de D par son aire[3] rapportée à celle de D. Calculer alors la probabilité cherchée.*
   (c) *Motiver la modélisation choisie.*
2.
   (a) *On choisit de coder une corde par l'angle au sommet qu'elle définit mesuré en degrés. Quel univers est-il alors pertinent de choisir ?*
   (b) *On définit la probabilité d'un intervalle de $[0, 180]$ par sa longueur[4] rapportée à celle de $[0, 180]$. Calculer alors la probabilité cherchée.*

[1] pour une notion usuelle, intuitive et utile de *probable* : qui va "sans doute" arriver

[2] Bien sûr, l'on pourra trouver des arguments pour motiver une modélisation privilégiée à l'exclusion des autres : mais de tels arguments sont du ressort de la *démarche scientifique* du mathématicien et non de la vérité interne à son champ de travail. (Pour le cas présent, la vraie question serait sans doute de savoir si le deuxième promeneur a *envie* de s'asseoir à côté du premier.)

[3] En toute rigueur, toutes les parties n'ont pas forcément une aire mais cette imprécision n'empêche en rien de répondre.

[4] En toute rigueur, toutes les parties n'ont pas forcément une longueur mais cette imprécision n'empêche en rien de répondre.

(c) *Motiver la modélisation choisie.*

3.

(a) *On choisit de coder une corde par la distance à O de son milieu. Quel univers doit-on alors choisir ?*

(b) *On définit la probabilité d'un intervalle de* $[0,1]$ *par sa longueur rapportée à celle de* $[0,1]$. *Calculer alors la probabilité cherchée.*

(c) *Motiver la modélisation choisie.*

4. *Commenter.*

Solution proposée.

1. (a) Puisque toute corde a son milieu dans le disque D, on doit choisir pour univers une *partie* de D. Réciproquement, montrons que tout point de D est le milieu d'une corde de C (*i. e.* tout point de D code une issue), ce qui montrera que l'univers est l'ensemble de *tous* les points de D, à savoir D tout entier.

(*Figure.* Soit $M \in D$. Soit D un diamètre passant par M. Appelons A et B les points d'intersection du cercle C avec la perpendiculaire à D passant par M. Alors la corde $[AB]$ a pour milieu M.)

(b) On se convaincra qu'une corde est de longueur supérieure ou égale à L ssi son milieu se situe dans le disque inscrit de T. Par conséquent, l'événement qui code "la corde tirée est de longueur supérieure ou égale à L" est le disque inscrit de T (appelons-le D'). Son aire vaut $\mathcal{A}(D') = \pi\left(\frac{1}{2}\right)^2 = \frac{\pi}{4}$. Or l'aire de D vaut $\mathcal{A}(D) = \pi(1)^2 = \pi$. On en déduit que la probabilité cherchée vaut $p(D') = \frac{\mathcal{A}(D')}{\mathcal{A}(D)} = \frac{\frac{\pi}{4}\cdot}{\pi} = \frac{1}{4}$.

(c) Pour engendrer une corde de façon aléatoire, on peut choisir "au hasard" son milieu dans le disque (à l'exception peut-être des cordes-diamètres correspondant aux milieux-centre). Le "au hasard" le plus habituel est le choix d'une mesure uniforme, ce qui permet au passage d'oublier les cordes-diamètres pathologiques (en effet, la probabilité de leur ensemble est alors la mesure d'un cercle rapportée à celle d'un disque, donc est nulle).

(a) Puisque toute corde définit un angle au centre compris entre $0°$ et $180°$ (cet angle vaut 0 quand les extrémités de la corde sont confondues et vaut $180°$ quand la corde est un diamètre), on doit choisir pour univers une *partie* de l'intervalle $[0,180]$. Réciproquement, il est clair que tout nombre α de $[0,180]$ définit une corde dont l'angle au centré vaut α degrés. L'univers est donc l'ensemble de *tous* les réels de $[0,180]$, à savoir l'intervalle $[0,180]$ tout entier.

(b) On se convaincra qu'une corde est de longueur supérieure ou égale à L ssi l'angle au centre qu'elle définit est compris entre $120°$ et $180°$. Par conséquent, l'événement qui code "la corde tirée est de longueur supérieure ou égale à L" est l'intervalle $[120,180]$. Sa longueur vaut $\ell([120,180]) = 180 - 120 = 60$. Or la longueur de $[0,180]$ vaut $180 - 0 = 180$. On en déduit que la probabilité cherchée vaut $p([120,180]) = \frac{\ell([120,180])}{\ell([0,180])} = \frac{60}{180} = \frac{1}{3}$.

(c) Pour engendrer une corde de façon aléatoire, on peut choisir "au hasard" ses extrémités sur le cercle. Pour des raisons d'isotropie (symétrie par rotation) validées par l'événement dont on cherche la probabilité, on n'a pas envie de distinguer les cordes sous-tendant des arcs de même longueur, ce qui motive l'oubli d'une des deux extrémités. Ensuite, une fois une première extrémité fixée, on peut choisir "au hasard" la seconde extrémité *le long du cercle*, choix et "hasard" naturellement codés par le cercle et la mesure uniforme sur ce dernier.

(a) Puisque le milieu d'une corde est un point de D, la distance de ce milieu an centre de D est inférieure ou égale à son rayon, *i. e.* la distance de ce milieu à O est inférieure ou égale à 1. Une distance étant par ailleurs toujours positive, on doit choisir pour univers une *partie* de l'intervalle $[0,1]$. Réciproquement, il est clair que tout nombre d de $[0,1]$ définit (au moins) une corde dont le milieu est à distance d du centre O (considérer les cordes tangentes au cercle de centre O et de rayon d). L'univers est donc l'ensemble de *tous* les réels de $[0,1]$, à savoir l'intervalle $[0,1]$ tout entier.

(b) Appartenir au disque inscrit dans T équivaut à être à distance de son centre inférieure ou égale à son rayon, donc équivaut à être à distance inférieure ou égale à $\frac{1}{2}$ du point O. Par conséquence, l'événement qui code "la corde tirée est de longueur supérieure ou égale à L" est l'intervalle $\left[\frac{1}{2},1\right]$. Sa longueur vaut $\ell\left(\left[\frac{1}{2},1\right]\right) = 1 - \frac{1}{2} = \frac{1}{2}$. Or la longueur de $[0,1]$ vaut $1 - 0 = 1$. On en déduit que la probabilité cherchée vaut $p\left(\left[\frac{1}{2},1\right]\right) = \frac{\ell\left(\left[\frac{1}{2},1\right]\right)}{\ell([0,1])} = \frac{\frac{1}{2}}{1} = \frac{1}{2}$.

(c) Il s'agit ici d'engendrer une corde de façon aléatoire en choisissant "au hasard" ses coordonnées polaires. L'isotropie du problème permet d'oublier l'argument : reste alors seul le module choisi "au hasard", d'où l'univers $[0,1]$ et la mesure uniforme sur ce dernier.

2. Les trois probabilités trouvées sont différentes, ce qui est paradoxal vu qu'elles désignent chacune "la probabilité" d'un même "événement" !

Nous sommes cependant trompés par *deux* abus de langage (dénoncés dans l'exercice 1) : d'une part, l'événement "la corde tirée a une longueur supérieure ou égale à L" est un événement du sens commun, d'autre part sa probabilité (le fait qu'il soit probable) est aussi du ressort du sens commun.

Pour évaluer cette probabilité (au sens commun) à l'aide de la mathématique, on commence *tout d'abord* par modéliser les issues de l'expérience "tirer une corde à extrémités dans C", *i. e.* par *choisir un univers*. Cela fait l'objet des questions a. L'événement (au sens commun) qui nous intéresse va alors être codé par un événement (mathématique), *i. e.* par une partie de l'univers.

Dans un second temps, on précise le sens de "au hasard" en *définissant une probabilité (mathématique)* sur l'univers choisi. Cela fait l'objet des questions b. Le nombre qui évalue alors la probabilité (au sens commun) de l'événement (au sens commun) qui nous intéresse vaut l'image de l'événement (mathématique) codant l'événement (au sens courant) qui nous intéresse par la probabilité (mathématique) choisie.

Il apparaît ainsi clairement que la probabilité (mathématique) de l'événement (au sens commun) qui nous intéresse dépend : d'une part du choix de l'univers (*i. e.* de la *modélisation des issues* de l'expérience), d'autre part du choix de la probabilité utilisée sur ce dernier (*i. e.* de la *modélisation de l'aléatoire* de l'expérience). Ces choix ont chacun été motivés lors des questions c.

3 Mise en jambe

1. *Soient $n \in \mathbf{N}^*$ et $A_1, ..., A_n$ des événements d'un espace probabilisé. Montrer la comparaison*

$$\mathbf{P}\left(\bigcap A_i\right) \geq \left(\sum \mathbf{P}(A_i)\right) - (n-1).$$

2. *On lance plusieurs dés équilibrés distinguables à six faces. Calculer la probabilité que la somme des nombres obtenus sur les faces du dessus soient paire.*

3. *Soient b, t et m trois naturels. On tire à l'aveugle t boules successivement et avec remise parmi b boules numérotées de 1 à b indistinguables au toucher. Quelle est la probabilité que la plus grande valeur tiré vaille m ? Estimer les chances que le plus grand numéro tiré vaille le nombre de boules tirées lorsque ce dernier est très grand.*

Solution proposée.

1. La comparaison voulue se réécrit $\mathbf{P}(\bigcap A_i) - 1 \geq (\sum \mathbf{P}(A_i)) - n$, ou encore $\mathbf{P'}(\bigcap A_i) \geq \sum \mathbf{P'}(A_i)$ où l'on a défini $\mathbf{P'} := \mathbf{P} - 1$. Une récurrence immédiate permettra alors de conclure si l'on montre le cas $n = 2$. Or ce dernier se réécrit $\mathbf{P}(A \cap B) \geq \mathbf{P}(A) + \mathbf{P}(B) - 1$, *i. e.* $1 \geq \mathbf{P}(A \cup B)$, ce qui est vrai. (Le cas $n = 1$ est trivial.)

2. On modélise chaque lancé par l'univers $\{1, ..., 6\}$ (car on s'intéresse à la valeur de la face du dessus) muni de la probabilité uniforme (car le dé est équilibré). Les dés étant distinguables, il est raisonnable de choisir pour univers de l'expérience le *produit* des univers précédents, à savoir $\{1, 2, ..., 6\}^n$ muni de la probabilité uniforme (où l'on a appelé n le nombre de dés). Notons (A_k) la suite des événements dont on cherche la probabilité.

Pour le cas d'un dé, il y a trois faces sur six qui montrent un nombre pair, d'où $\mathbf{P}(A_1) = \frac{3}{6} = \frac{1}{2}$.

Ensuite, si $n \geq 2$, la somme cherchée est paire ssi la somme sur les $n-1$ premiers lancé et la face du n-ième lancé ont même parité. En distinguant les deux parités, on obtient $\mathbf{P}(A_n) = \mathbf{P}(A_{n-1})\mathbf{P}(A_1) + \mathbf{P}({}^cA_{n-1})\mathbf{P}({}^cA_1)$, ce qui incite à raisonner par récurrence. L'étude des premiers cas nous permet d'intuiter $\forall k \in \mathbf{N}^*,\ \mathbf{P}(A_k) = \frac{1}{2}$, ce qui passe très bien du rang $n-1$ au rang n vu les égalités

$$\mathbf{P}(A_n) = \mathbf{P}(A_{n-1})\mathbf{P}(A_1) + \mathbf{P}({}^cA_{n-1})\mathbf{P}({}^cA_1) = \frac{1}{2}\frac{1}{2} + \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{2}\right) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

3. Le tirage étant *successif* et avec *remise*, on doit modéliser les issues par des listes *ordonnées* avec *répétitions possibles*. On choisit donc pour univers la puissance $\{1,2,...,b\}^t$. Les boules étant indistinguables au toucher et le tirage étant à l'aveugle, on doit supposer équiprobabilité sur *chaque* tirage, ce qui se propage à l'*ensemble* des tirages : on choisit donc pour probabilité la mesure uniforme sur $\{1,2,...,b\}^t$.

   Si $m > b$, l'événement est irréalisable et sa probabilité est nulle. On supposera donc $m \leq b$.

   Les issues nous intéressant étant les suites de $\{1,2,...,m\}$ qui atteignent m au moins une fois, l'événement correspondant est[5] $\{1,2,...,m\}^t \setminus \{1,2,...,m-1\}^t$, d'où son cardinal $m^t - (m-1)^t$. La probabilité cherchée vaut donc

$$\frac{m^t - (m-1)^t}{m^t} = 1 - \left(1 - \frac{1}{m}\right)^t.$$

Lorsque $t = m$ est très grand, le réel ci-dessus $1 - \left(1 - \frac{1}{m}\right)^m$ tend vers $1 - \frac{1}{e} \simeq 0,63$.

4 Paradoxe de Monty Hall

Un joueur a trois cartes devant lui, faces cachées (et indistinguables), une unique carte lui assurant le gain. Le joueur commence par désigner une carte (sans y toucher) : un maître du jeu (pouvant reconnaître les cartes faces cachées) retourne alors face visible une carte perdante qui n'est pas celle désignée par le joueur. Ce dernier choisit alors de retourner une carte parmi les deux restantes : cette carte lui assure le gain ou la perte.

Donner une stratégie permettant de gagner plus souvent qu'en jouant "au hasard". Commenter.

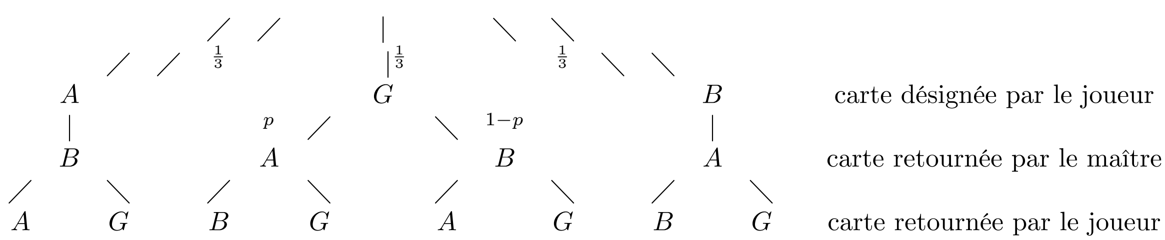
Solution proposée.

À la dernière étape, il reste deux cartes, l'une gagnante, l'autre perdante. L'indistinguabilité des cartes paraît nous imposer de choisir l'équiprobabilité du gain, d'où semble émerger un paradoxe dans l'énoncé : comment faire mieux qu'une chance sur deux de gagner sans information ? C'est que l'on ne cherche pas l'information au bon endroit : on ne peut certes pas distinguer les deux dernières cartes mais *on sait* que la carte désignée au départ est plus souvent perdante[6] (deux chances sur trois) – il convient donc de retourner l'autre carte afin de littéralement retourner la répartition des chances de gain/perte.

Formalisons tout cela.

Selon la stratégie choisie, on a affaire à des expériences différentes. Chacune d'entre elle possède un événement "le joueur gagne" dont on cherche à comparer les probabilités.

Dans tous les cas, choisissons une lettre G codant la carte gagnante et deux lettres A et B codant les cartes perdantes. Un jeu peut alors se représenter par l'arbre suivant :



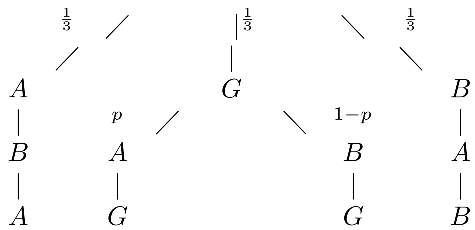
(les probabilités $\frac{1}{3}$ viennent du caractère indistinguables des cartes au premier "tour", caractère qui nous impose de choisir l'équiprobabilité ; les probabilités p et $1-p$ viennent de ce qu'on ne sait rien du coup du maître du jeu). L'univers est donc (selon ce codage) un ensemble de triplets à valeurs dans $\{A,B,G\}$. La mesure de probabilité dépendra de la stratégie du joueur à la dernière étape.

Dans le cas où le joueur joue "au hasard" (cette expression on ne peut plus ambigüe sous-entend "selon équiprobabilité"), on peut remplir toutes les arêtes de la dernière étape par $\frac{1}{2}$ et évaluer la probabilité cherchée : on trouve aisément $\frac{1}{2}$.

[5] si on avait oublié d'exclure le cas $m > b$, il aurait fallu remplacer m et $m-1$ resp. par $\min\{m,b\}$ et $\min\{m-1,b\}$, valeur égales (à b) dans le cas $m > b$: l'événement étudié aurait alors été vide, ce qui est cohérent.

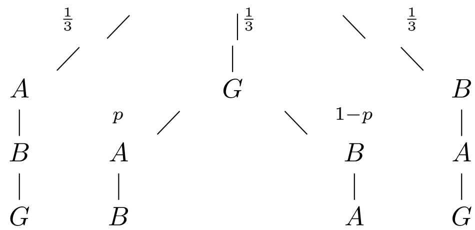
[6] On pourra mieux s'en convaincre en modifiant ainsi le jeu : on rajoute plein de cartes perdantes, on en désigne toujours une et le maître du jeu en garde toujours deux faces cachées dont la gagnante. Il est clair qu'en retournant la carte désignée (qui est presque toujours perdante) on perd presque à coup sûr.

Dans le cas où le joueur retourne la carte qu'il a désignée, l'arbre se simplifie en



et la probabilité cherchée vaut $\frac{1}{3}$.

Dans le cas où le joueur retourne la carte qu'il n'a pas désignée, l'arbre devient



et la probabilité cherchée $\frac{2}{3}$.

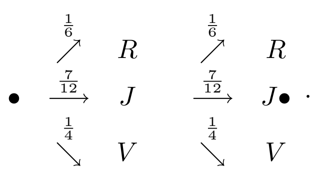
5 Un jeu de feux

Deux joueurs jouent avec un feu de signalisation déréglé. Toutes les cinq secondes, il change (peut-être) de couleur, prenant l'un des trois couleurs vert jaune et rouge. À son tour de jeu, le joueur perd si le feu passe au rouge, gagne si le feu vire au vert et passe sinon son tour à l'autre joueur. Le jeu s'arête dès qu'un joueur gagne ou perd. On suppose qu'à chaque tour la probabilité que le feu vire au vert vaut $\frac{1}{4}$ et est $1,5$ fois plus grande que celle de virer au rouge.

L'avantage est-il au premier joueur ou au second ?

Solution proposée.

Version arbolaire. Aidons-nous d'un arbre pour comprendre le jeu. Calculons auparavant la probabilité que le feu vire au rouge ($\frac{\text{proba de virer au vert}}{1{,}5} = \frac{\frac{1}{4}}{\frac{3}{2}} = \frac{1}{6}$) puis celle de virer au jaune par complémentaire à 1 de celles de virer au vert ou au rouge ($1 - \frac{1}{4} - \frac{1}{6} = \frac{12-3-2}{12} = \frac{7}{12}$) :



Bien observer que le hasard
d'une partie est inchangé
après deux feux jaunes successifs.

Notons p la probabilité que le premier joueur gagne. On lit sur l'arbre la relation $p = \frac{1}{4} + \frac{7}{12}\frac{1}{6} + \frac{7}{12}\frac{7}{12}p$. Isolant p et multipliant par 12^2, il vient $p\,(144-49) = 3\cdot 12 + 7\cdot 2$, soit $95p = 50$, d'où $p = \frac{5\cdot 10}{5\cdot 19} = \frac{10}{19}$. Le second joueur aura donc une probabilité de gain de $1 - \frac{10}{19} = \frac{9}{19}$, laquelle est inférieure à celle du premier joueur : l'avantage est donc au *premier* joueur.

Cela se comprendra très bien si l'on remplace les probabilités $\left(\frac{1}{6} < \frac{1}{4}\right)$ par $(0 < 1)$ et si l'on remplace le jeu par une roulette russe : même si les chances de gain/perte sont échangées à chaque tour, il vaut mieux *commencer* avec les plus grandes chances de gagner.

Version formalisée. On peut modéliser une partie de jeu par une suite finie de lettres parmi $\{V, J, R\}$ où la couleur prise par le feu est codée par son initiale. D'après les règles du jeu, cette suite doit finir par V ou R et ne comporter avant que des J. Une telle suite sera abrégée J^nV ou J^nR où n est le naturel comptant le nombre de lettres J. Ainsi l'événement codant "le premier joueur gagne" est-il

$$\left\{V, JR, J^2V, J^3R, J^4V, J^5R, ...\right\} = \coprod_{n\in\mathbf{N}} \left\{J^{2n}V, J^{2n+1}R\right\}.$$

Notons Ω l'ensemble de ces suites. (On peut éventuellement rajouter la suite vide infinie $J^\infty := (J, J, J, J, ...)$ où le jeu ne s'arrête pas.) Pour tout naturel k, notons J^k* l'événement codant "le feu est resté jaune pendant les n premiers tours".

Soit $\mathbf{P}$ une mesure de probabilité sur Ω. Pour tout naturel n, posons $\begin{pmatrix} v_{n+1} \\ j_{n+1} \\ r_{n+1} \end{pmatrix} := \begin{pmatrix} \mathbf{P}_{J^n*}\left(\{J^n V\}\right) \\ \mathbf{P}_{J^n*}\left(J^{n+1}*\right) \\ \mathbf{P}_{J^n*}\left(\{J^n R\}\right) \end{pmatrix}$: ainsi c_n dénote-t-il la probabilité que le feu vire à la couleur c au n-ième tour, les trois possibilités du $(n+1)$-ième tour aboutissant à la partition $J^n* = \{J^n V\} \amalg J^{n+1}* \amalg \{J^n R\}$. Supposons que $\mathbf{P}$ modélise bien les hypothèses. On a alors (à $n \in \mathbf{N}^*$ fixé)

$$v_n = \frac{1}{4}, \text{ d'où } r_n = \frac{v_n}{1,5} = \frac{\frac{1}{4}}{\frac{3}{2}} = \frac{1}{6} \text{ et } j_n = 1 - v_n - r_n = \frac{12-3-2}{12} = \frac{7}{12}.$$

Il en résulte (par conditionnement successif, à $N \in \mathbf{N}$ fixé)

$$\begin{aligned}
\mathbf{P}\left(J^N*\right) &= \mathbf{P}\left(J^0*\right) \times \mathbf{P}_{J^0*}\left(J^1*\right) \times \mathbf{P}_{J^1*}\left(J^2*\right) \times \cdots \times \times \mathbf{P}_{J^{N-1}*}\left(J^N*\right) \\
&= \mathbf{P}(\Omega) \times j_1 \times j_2 \times \cdots \times j_N = \left(\frac{7}{12}\right)^N, \text{ d'où} \\
\mathbf{P}\left(\left\{J^N V\right\}\right) &= \mathbf{P}\left(J^N*\right) \times \mathbf{P}_{J^N*}\left(\left\{J^N V\right\}\right) = \left(\frac{7}{12}\right)^N v_{N+1} = \left(\frac{7}{12}\right)^N \frac{1}{4} \text{ et} \\
\mathbf{P}\left(\left\{J^N R\right\}\right) &= \mathbf{P}\left(J^N*\right) \times \mathbf{P}_{J^N*}\left(\left\{J^N R\right\}\right) = \left(\frac{7}{12}\right)^N r_{N+1} = \left(\frac{7}{12}\right)^N \frac{1}{6}.
\end{aligned}$$

On en déduit la probabilité cherchée :

$$\begin{aligned}
\mathbf{P}\left(\coprod_{n\in\mathbf{N}} \left\{J^{2n}V, J^{2n+1}R\right\}\right) &= \sum_{n\in\mathbf{N}} \mathbf{P}\left(\left\{J^{2n}V, J^{2n+1}R\right\}\right) = \sum_{n\geq 0} \mathbf{P}\left(\left\{J^{2n}V\right\}\right) + \mathbf{P}\left(\left\{J^{2n+1}R\right\}\right) \\
&= \sum_{n\geq 0} \left(\frac{7}{12}\right)^{2n} \frac{1}{4} + \sum_{n\geq 0} \left(\frac{7}{12}\right)^{2n+1} \frac{1}{6} = \frac{1}{4} \frac{1}{1-\left(\frac{7}{12}\right)^2} + \frac{1}{6}\frac{7}{12}\frac{1}{1-\left(\frac{7}{12}\right)^2} \\
&= \frac{18+7}{72} \frac{144}{144-49} = \frac{25}{72}\frac{144}{95} = \frac{5}{1}\frac{2}{19} = \frac{10}{19}.
\end{aligned}$$

La même conclusion en découle : c'est le premier joueur qui est avantagé.

Remarque. Le formalisme n'a ici visiblement apporté aucune compréhension supplémentaire. Il permet cependant de justifier proprement (bien que lourdement et seulement *a posteriori*) le fait que le hasard du jeu est inchangé après deux feux premiers jaunes successifs, fait qui se formalise en $\forall A \subset \Omega,\ \mathbf{P}(A) = \mathbf{P}_{J^2*}\left(A \cap J^2*\right)$ et dont la vérité découle d'un calcul de sommes géométriques.

6 Probabilités et groupes non abéliens

On note p la probabilité que deux éléments dans un groupe fini donné commutent (les éléments sont tirés au hasard selon la loi uniforme). Montrer l'implication $p > \frac{5}{8} \implies p = 1$. Peut-on remplacer $\frac{5}{8}$ par une valeur inférieure ?

Notons G notre groupe fini : notre probabilité p s'exprime alors par

$$p = \frac{1}{|G|^2} \sum_{a\in G} \#\{b \in G\ ;\ ab = ba\}.$$

La cardinal dans la somme est aisé à évaluer lorsque a dans le centre Z de G : il vaut alors $|G|$. Dans le cas contraire, le commutant $\#\{b \in G\ ;\ ab = ba\}$ est un sous-groupe strict, donc est d'ordre $\leq \frac{|G|}{2}$. En séparant la sommation sur Z de la sommation ailleurs, on obtient l'inégalité

$$|G|^2 p \leq |Z|\,|G| + (|G| - |Z|)\frac{|G|}{2},\ i.\ e.\ p \leq \frac{|Z|}{2\,|G|} + \frac{1}{2}.$$

Si (par contraposée) on a $p < 1$, alors notre groupe G n'est pas commutatif, donc le quotient par son centre n'est pas cyclique (*cf.* exo précédent), donc le quotient ${}^{G}\!\diagup_{Z}$ n'est pas d'ordre premier, donc est d'ordre ≥ 4, ce qui s'écrit $\frac{|Z|}{|G|} \leq \frac{1}{4}$; en réinjectant dans l'inégalité ci-dessus, on obtient $p \leq \frac{1}{2}\frac{1}{4} + \frac{1}{2} = \frac{5}{8}$, *c. q. f. d.*.

Cherchons à réaliser l'égalité $p = \frac{5}{8}$, ce qui montrera que la valeur $\frac{5}{8}$ est minimale. Il s'agit de réaliser les égalités dans toutes les comparaisons précédentes, *i. e.* de trouver un groupe dont le centre soit d'indice 4 et dont tous les commutants soient d'indice 1 ou 2. Essayons un groupe d'ordre 8 (pour avoir le dénominateur de p) non abélien. À part deux produits semi-directs, il y a le groupe $\mathbf{H}_8$ des quaternions formé de la partie $\{\pm 1, \pm i, \pm j, \pm k\}$ de l'algèbre $\mathbf{Z}[i, j, k]$ quotientée par les relations $i^2 = j^2 = k^2 = -1$ et $ijk = 1$. On vérifiera que le centre de $\mathbf{H}_8$, qui vaut $\{\pm 1\}$, est bien d'indice 4 et que le commutant de $\pm i$, qui vaut $\{\pm 1, \pm i\}$, est bien d'indice 2 (de même pour $\pm j$ et $\pm k$).

7 Une comparaison

1. *Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants. On note $p_n := \mathbf{P}(A_n)$ pour tout entier $n \geq 1$. Montrer les comparaisons*
$$\forall k, n \in \mathbf{N},\ \mathbf{P}\left(\sum_{i=1}^{n} \mathbf{1}_{A_i} \geq k\right) \leq \frac{\left(\sum_{n\geq 1} p_n\right)^k}{k!}.$$
2. *Un joueur d'échecs affronte un ordinateur. Ils jouent une partie par jour jusqu'à ce que le joueur décide d'arrêter. Le niveau de l'ordinateur monte jour après jour (et l'attention du joueur diminue) : on admettra pour conséquence que les chances de gain de l'ordinateur augmentent chaque jour de 20%. Montrer que la probabilité que le joueur gagne au moins vingt parties est inférieure à un pour mille.*

Solution proposée.

1. Fixons n et k. On note E_k^n l'événement entre parenthèses. Soit $\omega \in E_k^n$. On a alors $\sum_{i=1}^{n} \mathbf{1}_{A_i}(\omega) \geq k$, donc la somme de gauche (qui ne contient que des 0 et des 1) contient au moins k termes 1, d'où k indices i tels que $\omega \in A_i$, ce qui montre l'inclusion $E_k^n \subset \bigcup_{I\subset\{1,\ldots,n\}}^{\operatorname{Card} I = k} \bigcap_{i\in I} A_i$. On en déduit
$$\begin{aligned}
\mathbf{P}(E_k^n) &\leq \mathbf{P}\left(\bigcup_{I\subset\{1,\ldots,n\}}^{\operatorname{Card} I=k} \bigcap_{i\in I} A_i\right) \leq \sum_{I\subset\{1,\ldots,n\}}^{\operatorname{Card} I=k} \mathbf{P}\left(\bigcap_{i\in I} A_i\right) \overset{\text{les } A_i \text{ sont}}{\underset{\text{indépendants}}{=}} \sum_{I\subset\{1,\ldots,n\}}^{\operatorname{Card} I=k} \prod_{i\in I} \mathbf{P}(A_i) \\
&= \sum_{I\subset\{1,\ldots,n\}}^{\operatorname{Card} I=k} \prod_{i\in I} p_i \overset{\substack{k!\text{ permutations}\\ \text{des éléments}}}{\underset{\substack{\text{d'une partie}\\ \text{à } k \text{ éléments}}}{=}} \sum_{J\in\{1,\ldots,n\}^k}^{J \text{ injective}} \frac{1}{k!} \prod_{i\in \operatorname{Im} J} p_j \overset{\text{tous les termes}}{\underset{\text{sont positifs}}{\geq}} \frac{1}{k!} \sum_{J\in\{1,\ldots,n\}^k} \prod_{i\in \operatorname{Im} J} p_j \\
&= \frac{1}{k!}\left(\sum_{i=1}^{n} p_i\right)^k \leq \frac{1}{k!}\left(\sum_{i=1}^{\infty} p_i\right)^k, \text{ c. q. f. d.}.
\end{aligned}$$
2. Notons A_n l'événement "le joueur gagne le n-ième jour". Ces événements sont indépendants (en négligeant les facteurs psychologiques résultant d'un gain ou d'une défaite) et leurs probabilités forment une suite géométrique de raison $\frac{1}{1+20\%} = \frac{5}{6}$. Leur somme valant $\sum_{n\geq 1} \mathbf{P}(A_1)\left(\frac{5}{6}\right)^n \leq \frac{5}{6}\frac{1}{1-\frac{5}{6}} = 5$, on en déduit au bout de N jours la majoration
$$\mathbf{P}\left(\sum_{i=1}^{N} \mathbf{1}_{A_i} \geq 20\right) \leq \frac{5^{20}}{20!} \simeq 3,92 \times 10^{-5} \leq 10^{-4},\ \text{c. q. f. d.}.$$

8 Remonter les arbres binaires : quatre variations bayésiennes et un paradoxe

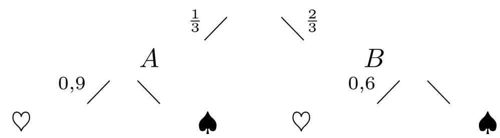
8.1 Cuisine

Alice apprend à Bob à faire des cannelés. Alice les réussit neuf fois sur dix et Bob six fois sur dix. Pour permettre à Bob de s'entraîner, Alice cuisine deux fois moins que Bob (ils ne cuisinent jamais ensemble).

On goûte de délicieux cannelés : quelle est la probabilité qu'ils soient fait de la main de Bob ?

Solution proposée.
L'expérience d'une cuisson peut-être codée un couple $(C, \clubsuit)$ où C code le cuisinier (Alice ou Bob) et où $\clubsuit$ code l'état final des cannelés (réussis ou ratés). On définit ainsi un univers Ω à quatre éléments. Définissons quatre événements : A code "Alice a cuisiné", B code "Bob a cuisiné", $\heartsuit$ "les cannelés sont délicieux" et $\spadesuit$ "les cannelés sont ratés" (observer les égalités $A \amalg B = \Omega = \heartsuit \amalg \spadesuit$).

Afin de mieux suivre les calculs à venir, on pourra s'aider de l'arbre suivant qui décrit une expérience de cuisson (les probabilités indiquées le long des arêtes seront précisées plus bas) :



Soit $\mathbf{P}$ une mesure de probabilité sur Ω répondant aux hypothèses. Ces dernières se traduisent successivement par $\mathbf{P}_A(\heartsuit) = \frac{9}{10}$, $\mathbf{P}_B(\heartsuit) = \frac{6}{10}$ et $\mathbf{P}(A) = \frac{1}{2}\mathbf{P}(B)$. Puisque $\mathbf{P}(A) + \mathbf{P}(B) = 1$, on en déduit $\mathbf{P}(A) = \frac{1}{3}$ et $\mathbf{P}(B) = \frac{2}{3}$, d'où l'on tire

$$\mathbf{P}(A \cap \heartsuit) = \mathbf{P}(A)\,\mathbf{P}_A(\heartsuit) = \frac{1}{3}\frac{9}{10} = \frac{3}{10} \qquad \text{et} \qquad \mathbf{P}(B \cap \heartsuit) = \mathbf{P}(B)\,\mathbf{P}_B(\heartsuit) = \frac{2}{3}\frac{6}{10} = \frac{4}{10}.$$

La probabilité cherchée vaut donc

$$\mathbf{P}_\heartsuit(B) = \frac{\mathbf{P}(B \cap \heartsuit)}{\mathbf{P}(\heartsuit)} = \frac{\mathbf{P}(B \cap \heartsuit)}{\mathbf{P}(A \cap \heartsuit) + \mathbf{P}(B \cap \heartsuit)} = \frac{1}{\frac{\mathbf{P}(A\cap\heartsuit)}{\mathbf{P}(B\cap\heartsuit)} + 1} = \frac{1}{\frac{3}{4} + 1} = \frac{4}{7}.$$

8.2 Dépistage

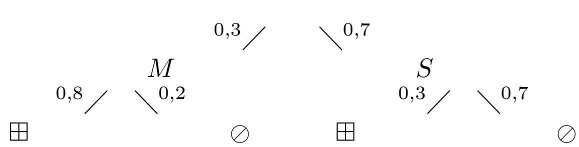
Afin de dépister une maladie affectant 30% d'une population, un test a été mis au point. On a observé statistiquement (à l'aide d'un autre test sûr) que : lorsqu'une personne est malade, le test le signale dans 80% des cas ; lorsqu'une personne est saine, le test le signale dans 70% des cas.

Discuter la fiabilité de ce test.

Solution proposée.
Reformulons la question : si le test nous dépiste, quelle est alors la probabilité d'être vraiment malade ? Autre question (apparemment la même) : si le test est négatif, quelle est alors la probabilité d'être vraiment sain ? Il s'agit donc d'évaluer des probabilités conditionnelles $\mathbf{P}_{\text{test positif}}$ (malade) et $\mathbf{P}_{\text{test négatif}}$ (sain).

L'expérience d'un test de dépistage sur une personne donnée peut-être codée un couple $(\ddagger, \circledast)$ où $\ddagger$ code l'état de santé de la personne (malade ou saine) et où $\circledast$ code le résultat du test (positif ou négatif). On définit ainsi un univers Ω à quatre éléments. Définissons quatre événements : M code "la personne est malade", S code "la personne est saine", $\boxplus$ "le test est positif" et $\oslash$ "le test est négatif" (observer les égalités $M \amalg S = \Omega = \boxplus \amalg \oslash$).

Afin de mieux suivre les calculs à venir, on pourra s'aider de l'arbre suivant qui décrit une expérience de test d'une personne donné (les probabilités indiquées le long des arêtes seront précisées plus bas) :



Soit $\mathbf{P}$ une mesure de probabilité sur Ω répondant aux hypothèses. Ces dernières se traduisent successivement par $\mathbf{P}(M) = 0,7$, $\mathbf{P}_M(\boxplus) = 0,8$ et $\mathbf{P}_S(\oslash) = 0,7$. Les autres probabilités de l'arbre s'en déduisent par complémentaire à 1, d'où l'on tire les probabilités des feuilles :

$$\begin{aligned}
\mathbf{P}(M \cap \boxplus) &= \mathbf{P}(M)\,\mathbf{P}_M(\boxplus) = 0,3 \cdot 0,8 = 0,24, \\
\mathbf{P}(M \cap \oslash) &= \mathbf{P}(M)\,\mathbf{P}_M(\oslash) = 0,3 \cdot 0,2 = 0,06, \\
\mathbf{P}(S \cap \boxplus) &= \mathbf{P}(S)\,\mathbf{P}_S(\boxplus) = 0,7 \cdot 0,3 = 0,21 \\
\text{et } \mathbf{P}(S \cap \oslash) &= \mathbf{P}(S)\,\mathbf{P}_S(\oslash) = 0,7 \cdot 0,7 = 0,49.
\end{aligned}$$

Les probabilités cherchées valent donc

$$\begin{aligned}
\mathbf{P}_{\boxplus}(M) &= \frac{\mathbf{P}(M\cap\boxplus)}{\mathbf{P}(\boxplus)} = \frac{\mathbf{P}(M\cap\boxplus)}{\mathbf{P}(M\cap\boxplus)+\mathbf{P}(S\cap\boxplus)} = \frac{1}{1+\frac{\mathbf{P}(S\cap\boxplus)}{\mathbf{P}(M\cap\boxplus)}} = \frac{1}{1+\frac{21}{24}} = \frac{24}{45} = \frac{8}{15} \simeq 53\% \\
\text{et } \mathbf{P}_{\oslash}(S) &= \frac{\mathbf{P}(S\cap\oslash)}{\mathbf{P}(\oslash)} = \frac{\mathbf{P}(S\cap\oslash)}{\mathbf{P}(M\cap\oslash)+\mathbf{P}(S\cap\oslash)} = \frac{1}{\frac{\mathbf{P}(M\cap\oslash)}{\mathbf{P}(S\cap\oslash)}+1} = \frac{1}{\frac{6}{49}+1} = \frac{49}{55} \simeq 89\%.
\end{aligned}$$

Ainsi, le test apparaît fiable (à 89%) lors d'un dépistage négatif mais ne l'est plus du tout (à peine plus d'une chance sur deux) en cas de dépistage positif – notons au passage que nos deux reformulations du problème n'étaient pas du tout équivalentes ! Par conséquent, le test sera utile si l'on souhaite repérer des sujets sains mais peu utile si l'on veut en repérer des malades.

8.3 Agriculture

Charles est agriculteur et souhaite, avant d'ensemencer son champ, séparer le bon grain de l'ivraie. Il préfère ne pas planter d'ivraie, quitte à jeter un peu de bon grain. Il s'aide pour cela d'une amie Denise qui se trompe dans 5% des cas quand elle reconnaît de l'ivraie et dans 10% des cas quand elle reconnaît un bon grain. Charles sait par ailleurs que 8% de sa semence est bonne à jeter.

Évaluer l'aide apportée par Denise.

Solution proposée.

Reformulons la question : si Denise reconnaît un bon grain, quelle est alors la probabilité que ce grain soit vraiment bon ? Autre question (apparemment la même) : si Denise reconnaît de l'ivraie, quelle est alors la probabilité que ce grain soit vraiment de l'ivraie ? Il s'agit donc d'évaluer des probabilités conditionnelles $\mathbf{P}_{\text{Denise voit bon grain}}$ (grain est bon) et $\mathbf{P}_{\text{Denise voit ivraie}}$ (grain est ivraie).

L'expérience de la reconnaissance d'un grain par Denise peut-être codée un couple $(G, \circledast)$ où G code le grain (bon ou ivraie) et où $\circledast$ code ce que voit Denise (bon ou ivraie). On définit ainsi un univers Ω à quatre éléments. Définissons quatre événements : B code "le grain est bon", I code "le grain est de l'ivraie",Db "Denise voit un grain bon" et Di "Denise voit de l'ivraie" (observer les égalités $B \amalg I = \Omega = Db \amalg Di$).

Afin de mieux suivre les calculs à venir, on pourra s'aider de l'arbre suivant qui décrit une expérience de reconnaissance de grain par Denise (les probabilités indiquées le long des arêtes seront précisées plus bas) : :

92% / \ 8%

B I

90% / \ 10% 5% / \ 95%

Db Di Db Di

Soit $\mathbf{P}$ une mesure de probabilité sur Ω répondant aux hypothèses. Ces dernières se traduisent successivement par $\mathbf{P}(I) = 8\%$, $\mathbf{P}_{Db}(I) = 10\%$ et $\mathbf{P}_{Di}(B) = 5\%$. Les autres probabilités de l'arbre s'en déduisent par complémentaire à 1, d'où l'on tire les probabilités des feuilles :

$$\begin{aligned}
\mathbf{P}(B\cap Db) &= \mathbf{P}(B)\,\mathbf{P}_B(Db) = 92\%\cdot 0,9 = 82,8\%, \\
\mathbf{P}(B\cap Di) &= \mathbf{P}(B)\,\mathbf{P}_B(Di) = 92\%\cdot 0,1 = 9,2\%, \\
\mathbf{P}(I\cap Db) &= \mathbf{P}(I)\,\mathbf{P}_I(Db) = 8\%\cdot 0,05 = 0,4\% \\
\text{et } \mathbf{P}(I\cap Di) &= \mathbf{P}(I)\,\mathbf{P}_I(Di) = 8\%\cdot 0,95 = 7,6\%.
\end{aligned}$$

Les probabilités cherchées valent donc

$$\begin{aligned}
\mathbf{P}_{Db}(B) &= \frac{\mathbf{P}(B\cap Db)}{\mathbf{P}(Db)} = \frac{\mathbf{P}(B\cap Db)}{\mathbf{P}(B\cap Db)+\mathbf{P}(I\cap Db)} = \frac{1}{1+\frac{\mathbf{P}(I\cap Db)}{\mathbf{P}(B\cap Db)}} = \frac{1}{1+\frac{4}{828}} = \frac{828}{832} = \frac{207}{208} \simeq 99,5\% \\
\text{et } \mathbf{P}_{Di}(I) &= \frac{\mathbf{P}(I\cap Di)}{\mathbf{P}(Di)} = \frac{\mathbf{P}(I\cap Di)}{\mathbf{P}(B\cap Di)+\mathbf{P}(I\cap Di)} = \frac{1}{\frac{\mathbf{P}(B\cap Di)}{\mathbf{P}(I\cap Di)}+1} = \frac{1}{\frac{92}{76}+1} = \frac{76}{168} = \frac{19}{42} \simeq 45,2\%.
\end{aligned}$$

Ainsi, Denise apparaît très utile (à plus de 99%) lors qu'elle reconnaît un bon grain mais ne l'est plus du tout (moins d'une chanche sur deux) quand elle reconnaît de l'ivraie. Ce dernier cas n'est cependant pas grave puisque Charles veut éliminer l'ivraie pour garder le bon grain : quitte à jeter un peu de bon grain, Denise lui sera très utile dans tous les cas.

8.4 Interrogatoires

Des enquêteurs interrogent des individus chacun suspecté d'un méfait. À l'issue de l'interrogatoire, chaque individu avoue ou non sa culpabilité. On note C (resp. I, resp. A) les événements "l'individu interrogé est coupable (resp. est innocent, resp. avoue)". On définit $p := \mathbf{P}(C)$ et $r := \frac{\mathbf{P}_I(A)}{\mathbf{P}_C(A)}$.

1. *Proposer un modèle pour cette expérience.*
2. *Interprêter le rapport r.*
3. *Donner une condition simple équivalant à $\mathbf{P}_A(C) > \mathbf{P}(C)$ et qui met en jeu r. Commenter.*

Solution proposé.

1. Les aveux sont observés *suite* aux interrogatoires ; l'innocence ou non d'un individu résultant par ailleurs d'un fait *passé*, il est naturel d'utiliser un modèle statistique. L'univers sera l'ensemble des individus et la probabilité d'un événement sa fréquence.
2. La probabilité $\mathbf{P}_I(A)$ mesure un degré de terreur de l'interrogatoire : que peut-on imaginer d'autre qui fasse avouer des innocents ? La probabilité $\mathbf{P}_C(A)$ mesure par ailleurs sa fiabilité : l'interrogatoire est une réussite (sur le plan de l'information extraite) si tous les coupables ont avoué. Ainsi le rapport r mesure-t-il le ratio terreur/fiabilité. Les enquêteurs cherchent idéalement à maximiser $\mathbf{P}_C(A)$ tout en minimisant $\mathbf{P}_I(A)$, ce qui a pour effet de minimiser r (idéalement nul).
3. On a $\mathbf{P}_A(C) = \frac{\mathbf{P}_C(A)}{\mathbf{P}(A)}\mathbf{P}(C) = \frac{\mathbf{P}_C(A)\ \mathbf{P}(C)}{\mathbf{P}(C)\mathbf{P}_C(A)+\mathbf{P}(I)\mathbf{P}_I(A)} = \frac{p}{p+(1-p)r}$, d'où les équivalences

$$\mathbf{P}_A(C) > \mathbf{P}(C) \iff \frac{p}{p+(1-p)r} > p \iff \frac{1}{p+(1-p)r} > 1 \iff 1 > p+(1-p)r \iff 1 > r.$$

   Ainsi, si les enquêteurs utilisent des moyens d'interrogation peu louables (torture...) au point de faire croître le ratio terreur/fiabilité au-delà de 1, la conséquence $\mathbf{P}_A(C) < \mathbf{P}(C)$ sera qu'un aveu *diminue* la culpabilité, dénuant par conséquent l'interrogatoire de sa fin. Sans surprise, la "rudesse" des enquêteurs doit être mesurée : trop faible, elle ne fait pas avouer des coupables, excessive, elle pousse des innocents à avouer.

8.5 Paradoxe de Simpson

On étudie chez une population l'efficacité d'un médicament. Pour cela, on la compare à celle d'un placebo. Les effectifs des caractères "guéri" et "non guéri" sont indiqués dans les tableaux suivants, répartis selon le sexe de l'individu et son traitement (médicament ou placebo) :

hommes	guéri	non guéri
médicament	36	24
placebo	14	6

femmes	guérie	non guérie
médicament	4	16
placebo	18	42

.

Est-ce que le médicament marche mieux que le placebo pour les hommes ? Pour les femmes ? Pour toute la population ? Commenter ces résultats et votre prise de décision.

Solution proposée.

Il est naturel de dire que le médicament marche mieux que le placébo si la fréquence de guérison parmi ceux qui ont pris le médicament est plus grande que celle de ceux qui ont pris le placebo. Calculons donc ces fréquences[7] :

	homme médicament	homme placebo	femme médicament	femme placebo	individu médicament	individu placebo
effectif total	60	20	20	60	80	80
effectif guéri	36	14	4	18	40	32
fréquence de guérison	$\frac{36}{60} = \frac{3}{5}$ $\underline{= 0,6}$	$\frac{14}{20} = \frac{7}{10}$ $\underline{= 0,7}$	$\frac{4}{20} = \frac{1}{5}$ $\underline{= 0,2}$	$\frac{18}{60} = \frac{3}{10}$ $\underline{= 0,3}$	$\frac{40}{80} = \frac{1}{2}$ $\underline{= 0,5}$	$\frac{32}{80} = \frac{2}{5}$ $\underline{= 0,4}$

[7] Attention à ne pas calculer la fréquence parmi les guéris des individus ayant pris le médicament (resp. placebo) ! Imaginer en effet que tout le monde a pris le médicament et que personne n'a guéri sauf un individu : cette fréquence vaudra alors 1 mais on ne pourra certainement pas dire que le médicament a été efficace.

On voit alors que le médicament marche moins bien que le placebo pour les hommes et pour les femmes mais mieux pour toute la population. Ainsi, après découpage de la population en deux, le placebo marche mieux que le médicament sur les deux sous-populations mais marche moins bien sur la population total, ce qui a l'air paradoxal[8]. Nous sommes appelés à critiquer au moins trois niveaux.

Confusion entre réalité & modèle. Cet aspect contre-intuitif vient du fait que nous surinterprétons l'expression "marcher mieux que". Nous avons choisi de la traduire mathématiquement par une comparaison de nombres qui modélisent l'efficacité (à savoir : les fréquences de guérison). Le calcul numérique donne ce qu'il donne : si son interprétation ne nous plaît pas, il faut alors changer notre définition de "marcher mieux que" (en avez-vous une meilleure ?) ou bien accepter humblement que notre intuition s'est égarée en faisant dire à la mathématique plus qu'elle ne le peut.

Choix de la population. À chacun de se placer dans la sous-population qui le concerne le plus. Si l'on exagère, quel intérêt à comparer les efficacités du médicament et du placebo sur des lamas ou sur des insectes et intégrer ces derniers dans la population ? Il faut raffiner sa population le plus possible (on pourrait peut-être préciser des tranches d'âges, des localisations géographiques, des prédispositions médicales...) tout en gardant suffisamment d'individus pour que les statistiques gardent un sens. Dans l'exercice, à chacun de comparer l'efficacité selon son sexe : on constate dans les deux cas que le placebo a mieux marché que le médicament, ce qui doit pousser à préférer celui-là à celui-ci. Cette comparaison n'exclut bien sûr pas une tierce alternative : on pourrait imaginer une autre partition de la population (par exemple selon la couleur des yeux) qui aboutisse à une conclusion opposée[9] à celle de la partition selon les sexes (il conviendra alors se restreindre à l'ntersection des population auxquelles on appartient).

Pertinence du protocole. Même si la moitié de la population a reçu le médicament et l'autre moitié le placebo (ce qui rend objective la comparaison globale des deux produits), il y a des différences significatives d'effectifs entre les sous-populations testées (trois fois plus de femmes que d'hommes ont reçu le placebo au lieu du médicament), ce qui casse l'objectivité des résultats pour chacun des sexes – ainsi que la décision qui en découlait au paragraphe précédent. Sans doute pourrions-nous imaginer une autre bipartition de la population, en effectifs égaux, que le choix de la population n'aura pas permis de prévoir : cela est même en un sens toujours possible[10] (bien qu'artificiel).

9 Chapeaux et devinette colorée

Des personnes portant chacune un chapeau se mettent en file indienne, chaque personne voyant les chapeaux de toutes les personnes devant elle (mais pas le sien). En partant depuis la queue de la file, chaque personne dit à voix haute (de sorte à être entendue par la personne devant) une couleur qui est censée être celle de son chapeau. On suppose connu de toutes le nombre de couleurs possiblement utilisées.

1. *Proposez à la file de personnes une stratégie qui permette de maximiser la probabilité de deviner toutes les bonnes couleurs. Que vaut cette probabilité maximale ?*
2. *Généraliser à un nombre infini de couleurs. On pourra montrer que, pour tout ensemble C infini, les ensembles $\mathbf{Z}^{(C)}$ et C sont équipotents.*
3. *Généraliser à un nombre infini de personnes. On pourra montrer d'une part que, pour tout ensemble C infini, les ensembles $\mathbf{Q}\left((X_c)_{c\in C}\right)$ et C sont équipotents, d'autre part que toute forme linéaire sur un sous-espace vectoriel se prolonge à tout l'espace vectoriel.*

Solution proposée.

1. Le chapeau de la n-ième personne peut être décrit comme les chapeaux vus par la n-ième personne desquels on a "retranché" ceux vu par la $(n-1)$-ième personne. Ainsi, si l'on trouve un moyen, à travers ce que dit une personne, de coder l'information qu'elle voit de manière "*additive*", la n-ième personne pourra alors deviner la couleur de son chapeau en retranchant à ce qu'elle voit ce que son voisin de derrière aura vu (à travers ce qu'il aura *dit*). Un moyen simple pour cela est de numéroter les couleurs $1, 2, 3..., c$ et d'additionner *modulo* c : une personne annoncera donc la somme des couleurs qu'elle voit devant elle. Seule la première personne pourrait ne pas deviner à coup sûr la couleur de son chapeau mais elle ne dispose de toute façon d'*aucune* information et ne pourra pas mieux faire qu'une chance sur c : autant que cette chance serve aux autres. La probabilité maximale vaut donc $\frac{1}{c}$ et est atteinte pour la stratégie sus-décrite.

[8] Un article de *Pour la science* (n°429 juillet 2013) est consacré à ce paradoxe.
[9] Cela est fait dans l'article sus-cité.
[10] On renvoie à l'article sus-cité pour plus de détails.

2. Notons C l'ensemble des couleurs. Pour continuer d'utiliser l'idée précédente, il suffit de munir C d'une structure de groupe additif. Par transport de structure, il suffit de trouver un groupe équipotent à C, ce qui est le cas[11] de $\mathbf{Z}^{(C)}$.

   (Il est tentant de restreindre C aux couleurs des chapeaux portés par les personnes, lesquelles sont en nombre fini, ce qui nous ramènerait au cas précédent ; mais comment transmettre cette information à toutes les personnes ?)

3. Notons P l'ensemble des personnes. Pour continuer d'utiliser l'idée de sommer les couleurs (ce qui pose problème quand on essaie d'en additionner une infinité), montrons qu'il suffit de prolonger la fonction $\sum_P$ définie sur $C^{(P)}$ en une fonction Σ additive sur C^P.

   Notons $c \in C^P$ la fonction qui à une personne associe la couleur de son chapeau. Pour toute personne $p \in P$, notons $c_p \in C^P$ la fonction qui est nulle partout sauf en p où elle vaut $c(p)$. Appelons π la première personne à parler : elle connaît $c - c_\pi$ et dira la couleur $\Sigma(c - c_\pi)$. Toute autre personne p à parler ensuite, connaissant par ailleurs $c - c_p$ ainsi que $c(\pi)$ (*a fortiori* $c - c_p - c_\pi$) peut retrancher de ce qu'a dit π (à savoir $\Sigma(c - c_\pi)$) sa propre connaissance filtrée par Σ (à savoir $\Sigma(c - c_p - c_\pi)$) et obtenir ainsi $\Sigma(c_p) = c(p)$.

   Pour pouvoir effectuer un tel prolongement, il est naturel de Zornifier l'ensemble des prolongements : cela marche bien car il est aisé de prolonger "à la limite". Le problème sera de prolonger "au successeur" : en effet, si φ est un morphisme de groupes défini sur un sous-groupe G strict de C^P et si f est une fonction hors de G, on pourra prolonger f à $G + \mathbf{Z}f$ si cette dernière somme est *directe* : mais sinon ? Pour pallier ce problème, on va remplacer la structure de groupes (et l'additivité de φ) par celle d'espaces vectoriels (et par la linéarité de φ) : un vecteur hors d'un sous-espace vectoriel G engendre en effet une droite qui est toujours en somme directe avec G. Pour ce faire, il suffit de munir C d'une structure de corps (la fonction $\sum_P$ sera alors une forme linéaire sur le sous-C-espace vectoriel $C^{(P)}$), ce qui peut être fait (par transport de structure) en lui trouvant un corps équipotent. Montrons que c'est le cas du corps $K_C := \mathbf{Q}\left((X_c)_{c \in C}\right)$. Notons A_C l'anneau $\mathbf{Q}\left[(X_c)_{c \in C}\right]$.

   Déjà, toute fraction rationnelle étant le quotient de deux polynômes, on a une surjection $A_C^2 \twoheadrightarrow K_C$; puisque A_C est infini (il contient $\mathbf{Q}$), il est équipotent à son carré, d'où une surjection $A_C \twoheadrightarrow K_C$, ce qui revient à une injection $K_C \hookrightarrow A_C$ (après choix d'une section). On a par ailleurs une injection $A_C \hookrightarrow K_C$ (induite par l'inclusion), ce qui montre par BBBCDSZ l'équipotence de K_C et A_C. Maintenant, A_C est la réunion disjointe sur les parties P finies de C des droites $\mathbf{Q}\prod_{c \in P} X_c$: le domaine de réunion étant équipotent[12] à C et chaque droite étant équipotente à $\mathbf{N}$, l'ensemble A_C est équipotent à C copies de $\mathbf{N}$, *i. e.* à $C \times \mathbf{N}$, donc à C.

Remarque. On peut rendre le jeu plus dramatique en menaçant de mort les personnes ne donnant pas la couleur de leur chapeau (et en promettant la vie sauve aux autres). L'utilisation plus ou moins cachée de l'axiome du choix dans les questions 2 et 3, permettant d'aboutir à un sauvetage optimal, devient alors capitale : comment en pratique utiliser efficacement un *axiome* (dont on ne peut prouver la vérité) dans une situation réelle de vie ou de mort ? Libre commentaire...

[11] C'est un résultat classique de théorie des ensembles que la "puissance" $F^{(E)}$ a pour cardinal le plus grand de ceux de E et F lorsque l'un de ces derniers est infini. On utilise au passage une version de l'axiome du choix : tout ensemble infini est équipotent à son carré.

[12] C'est un résultat classique de théorie des ensembles : l'ensemble des parties finies d'un ensemble infini I est équipotent à I.