Des familles de deux enfants

Claudine Schwartz, IREM de Grenoble
Professeur, Université Joseph Fourier

Les questions 1 et 4 sont posées dans le dernier numéro de « Pour la Science » (n°336, octobre 2005, article de JP Delahaye, page 92-94). Ce texte reprend les questions de cet article, en les réorganisant et en commentant quelques résultats.

1) Une famille a 2 enfants, dont une fille. Quelle est la probabilité p_1 que l'autre enfant soit une fille ?

2) On va dans la maison d'une famille de 2 enfants. C'est une fille qui ouvre la porte. Quelle est la probabilité p_2 que l'autre enfant soit une fille ?

3) Une famille a 2 enfants, dont une fille née un premier septembre. Quelle est la probabilité p_3 que l'autre enfant soit une fille ?

4) Une famille a 2 enfants, dont une fille qui se nomme Sophie. Quelle est la probabilité p_4 que l'autre enfant soit une fille ?

5) Une famille a 2 enfants, dont un qui se nomme Dominique. Quelle est la probabilité p_5 que l'autre enfant soit une fille ?

6) Une famille a 2 enfants, dont un enfant qui se nomme Dominique. Quelle est la probabilité p_6 que cette famille ait deux filles ?

Selon une enquête micro-trottoir, les biologistes et les physiciens refusent de répondre à ces questions : ils ne voient pas où on veut en venir, la question leur paraît trop absurde.

Les mathématiciens acceptent plus fréquemment de répondre, soit en donnant des réponses immédiates (et c'est souvent la même valeur pour toutes les questions posées), soit en donnant des réponses qui, sous des présentations diverses, sont celles proposées ci-dessous.

Probabilistes ou non, nous sommes beaucoup à penser que ces questions sont de nature à en écarter plus d'un des probabilités, surtout si celui qui les pose finit par imposer une solution comme étant « la solution juste ».

On demande ici « quelle est la probabilité… » : c'est donc qu'on traite d'un modèle ou d'une expérience aléatoire et pas d'une seule famille à 2 enfants vivant sur une île déserte. Plusieurs expériences et modèles peuvent alors être proposés, d'où plusieurs réponses envisageables.

On choisit, sauf mention contraire dans la partie 4-2, **les contraintes suivantes : le sexe du premier et du deuxième enfant sont indépendants et à chaque naissance, garçons et filles sont équiprobables**.

1) Deux enfants dont une fille.

Distinguons l'aîné et le second enfant.

On note (F,G) une famille dont l'aîné est une fille et le cadet un garçon.
Les quatre possibilités (F,F), (F,G), (G,F) et (F,F) étant toutes de probabilité 1/4, on peut écrire :

$$p_1 = \text{Prob } [\ (\text{F,F}) \ / \ \{(\text{F,F}),(\text{F,G}),(\text{G,F})\}\ ] = \frac{1/4}{3/4} = \frac{1}{3}$$

On a donc une chance sur 3 que l'autre enfant soit une fille.

2) Une fille ouvre la porte.

2-1. Etre accueilli par une fille prouve qu'il y a au moins une fille dans cette famille. Si on en reste là, on est ramené à la situation de la question 1. Mais on peut aussi se poser la question de la règle d'ouverture de la porte par un des deux enfants. D'où une nouvelle étude possible.

2-2. On note OF l'événement « une fille ouvre la porte ».
Soit α la probabilité de OF, sachant que la famille comporte une fille et un garçon.

Les probabilités qui nous intéressent sont :

$$\text{Prob } ((\text{F,G}) \text{ et OF}\ ) = \alpha/4 \qquad \text{Prob } ((\text{G,F}) \text{ et OF}) = \alpha/4$$

$$\text{Prob } ((\text{F,F})) = \text{Prob } ((\text{F,F}) \text{ et OF}) = 1/4$$

$$p_2 = \text{Prob } [\ \{(\text{F,F})\} \ / \ \{\{(\text{F,F}),(\text{F,G}),(\text{G,F})\} \text{ et OF}\}\ ] =$$

$$\text{Soit } p_2 = \frac{1}{1+2\alpha}.$$

Le cas 2-1 coïncide avec α = 1, c'est-à-dire au cas où, s'il n'y a qu'une seule fille, c'est elle qui ouvre. Si on ne sait rien des règles d'ouverture de porte, on peut prendre α = 0,5, et la probabilité cherchée vaut 0,5.

2-3. On pourrait aussi paramétrer en notant β la probabilité que ce soit l'aîné qui ouvre :

$$\text{Prob } ((\text{F,G}) \text{ et OF}) = \beta/4 \qquad \text{Prob } ((\text{G,F}) \text{ et OF}) = (1-\beta)/4$$

$$\text{Prob } ((\text{F,F}) \text{ et OF}) = 1/4$$

Dans ce cas, la probabilité que l'autre enfant soit une fille est :

$$p_2' = \text{Prob } [\ \{(\text{F,F}) \text{ et OF}\} \ / \ \{\{(\text{F,F}),(\text{F,G}),(\text{G,F})\} \text{ et OF}\}\ ] = 1/2$$

La probabilité que l'autre enfant soit une fille vaut ici 1/2, quelle que soit la valeur de β.

3) Une fille née un premier septembre.

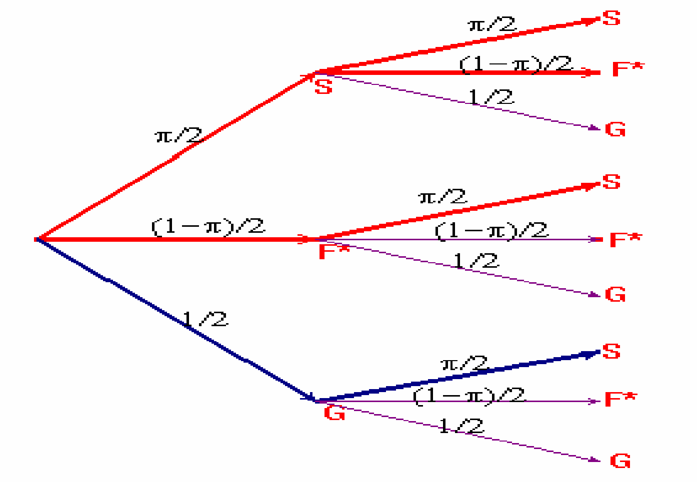
Soit π la probabilité qu'une fille soit née un premier septembre.
Soient :
- S l'événement « une fille née un 1^er septembre » ;
- F* l'événement « une fille qui n'est pas née un premier septembre ».

On a (voir arbre de probabilité ci-dessous) :

$$\text{Prob}((S,G)) = \pi/4$$

$$\text{Prob}((G,S)) = \pi/4$$

$$\text{Prob}((S,F)) = \text{Prob}(\{(S,S),(S,F^*)\}) = \pi/4 \text{ et } \text{Prob}((F^*,S)) = (1-\pi)\pi/4$$



D'où :

$$p_3 = \text{Prob}\left[\ \{(F^*,S),(S,F)\}\ /\ \{(F^*,S),(S,F),(S,G),(G,S)\}\ \right] = \frac{2-\pi}{4-\pi}$$

On remarque que :

Comme la probabilité de naître un premier septembre est petite, p_3 est proche de 1/2.

4) Une fille qui se nomme Sophie.

4-1. Dans l'article de Delahaye de « pour la Science », il est écrit :
« Notons E_1 l'enfant qui s'appelle Sophie (on peut introduire cette notation puisque cet enfant est unique) et E_2 l'enfant qui ne s'appelle pas Sophie. Deux cas sont possibles, et de même probabilité : E_2 est un garçon, E_2 est une fille. Donc p_4 = 1/2. »

Détaillons cet argument.

Notons :
- A l'événement « l'enfant qui ne s'appelle pas Sophie est une fille » ;
- S_1 l'événement « l'aîné des deux enfants est une fille qui s'appelle Sophie » ;
- S_2 l'événement « le plus jeune des deux enfants est une fille qui s'appelle Sophie ».

On suppose que deux enfants ne peuvent avoir le même prénom ; on peut écrire :

$$\text{Prob } (S_1 \text{ ou } S_2) = \text{Prob } (S_1) + \text{Prob } (S_2)$$

$$\text{Prob } ((A \text{ et } S_1) \text{ ou } (A\text{et } S_2)) = \text{Prob } (A \text{ et } S_1) + \text{Prob } (A \text{ et } S_2)$$

On a : $p_4 = \text{Prob } (A / (S_1 \text{ ou } S_2) = [\text{ Prob } (A \text{ et } S_1) + \text{Prob}(A \text{ et } S_2)\ ] / \text{Prob } (S_1 \text{ ou } S_2)$

La probabilité que le second enfant soit une fille, sachant que l'aîné est une fille, ou alors sachant qu'elle s'appelle Sophie (ou Dominique, voir question 5), est 0,5.

D'où :

$$\text{Prob } (A \text{ et } S_1) = \text{Prob } (A / S_1) \times \text{Prob } (S_1) = 0{,}5 \times \text{Prob } (S_1)$$

De même : $\text{Prob } (A \text{ et } S_2) = \text{Prob } (A / S_2) \times \text{Prob } (S_2) = 0{,}5 \times \text{Prob } (S_2)$

Finalement : $\boldsymbol{p_4 = 1/2}$

La seule hypothèse dont on se sert ici à propos du prénom est que deux enfants de la même famille n'ont jamais le même prénom.

4-2. Delahaye propose une expérience pour mieux comprendre : on tire au hasard deux cartons parmi 100, dont 50 sont étiquetés G et 50 sont étiquetés F, et parmi les cartons F, un est étiqueté S (**avec une telle expérience, on oublie qui est l'aîné des enfants**).

Cette expérience revient à considérer un modèle dans lequel la probabilité, pour une famille de 2 enfants, d'avoir une Sophie est $\dfrac{99}{(100\times 99)/2} = 0{,}02$. La probabilité d'avoir deux filles dont une Sophie est $\dfrac{49}{(100\times 99/2)}$ et la probabilité d'avoir deux filles sachant qu'on a une Sophie est :

$$p_4' = 49/99 \approx 0{,}495$$

La probabilité est voisine de 0,5.

La répartition des compositions des familles avec ce modèle vérifie :

$$\text{Prob (deux filles)} = \text{Prob (deux garçons)} = (50\times 49/2)/(100\times 99/2) = 49/198 \approx 0{,}247$$

$$\text{Prob (une fille et un garçon)} = 50\times 50/(100\times 99/2) = 50/99 \approx 0{,}505.$$

La question 1 traitée avec ce modèle donne :

$$p_1' = (25\times 49)/(25\times 49+50\times 50) = 49/149.$$

Cette probabilité p_1' est proche de 1/3.

Ce modèle est tout à fait acceptable pour la composition des familles ; il n'est pas plus éloigné de la réalité que celui qu'on a considéré (car l'égalité des sexes à la naissance n'est pas tout à fait respectée, le modèle avec 105 garçons pour 100 filles est plus pertinent que l'équiprobabilité des sexes à la naissance). Le résultat trouvé dépend du nombre de cartons. Regardons ce qui se passe si on augmente le nombre de cartons.

Pour N cartons G et N cartons F, la probabilité d'avoir deux cartons F sachant qu'on a au moins un carton F est :

$$(N\text{-}1)/(3N\text{-}1)$$

Elle tend vers 1/3 quand N tend vers l'infini.

La probabilité d'avoir deux filles est $(N\text{-}1)/2(2N\text{-}1)$ et tend vers 1/4 lorsque N tend vers l'infini, ainsi que celle d'avoir deux garçons.

Si parmi les N cartons F, un seul est étiqueté Sophie, (ce qui exclut la possibilité d'avoir deux fois le prénom de Sophie), la probabilité d'avoir deux filles sachant qu'on a une Sophie est :

$$(N\text{-}1)/(2N\text{-}1)$$

Quand N tend vers l'infini, elle tend vers 1/2, mais faire tendre N vers l'infini n'est pas pertinent ici, car la probabilité qu'une fille se nomme Sophie est $1/N$ et tend alors vers 0.

Dans l'article de « Pour la Science », Delahaye propose de simuler l'expérience des tirages de cartons. Sur un grand nombre de simulations, on trouvera, parmi les familles avec une Sophie, une fréquence voisine de 0,5 pour les familles de deux filles. On peut alors construire des intervalles de confiance pour la probabilité p'_4 ; on ne trouvera pas la valeur exacte de p'_4, ni celle de p_4, mais on saura qu'elles ne sont pas 1/3

5) Un enfant qui se nomme Dominique.

L'information apportée par un prénom est utilisée pour distinguer un enfant des autres. Pour le sexe de l'autre enfant, peu importe que ce prénom soit rare ou non, que ce soit celui d'une fille ou d'un garçon.
On peut faire le même calcul que dans la partie 4-1, dans lequel le fait que Sophie soit une fille n'intervient pas. Donc, s'il y a un enfant qui s'appelle Dominique, la probabilité que l'autre soit une fille est aussi 1/2.
Soit :

$$\boldsymbol{p_5 = 0{,}5}$$

En fait, pour mieux comprendre que dans la question 1 on trouve 1/3 alors qu'en rajoutant la connaissance d'un prénom on trouve 1/2, on peut imaginer le modèle suivant :

On a k particules, numérotées de 1 à k, qui peuvent être dans l'état codé 1 ou dans l'état codé 2, avec équiprobabilité de chaque état et indépendance des états de chacune d'elles. Ces particules sont par ailleurs placées dans des boîtes : il y a N boîtes, $N > k$; une boîte contient au plus une particule. Le placement des particules est indépendant de leur état. La loi des placements est une loi de probabilité, équirépartie ou non, sur l'ensemble E des sous-ensembles de k boîtes parmi N.

(a) Quelle est la probabilité, sachant qu'il y a au moins une particule dans l'état 1, que toutes les autres soient aussi dans l'état 1 ?

(b) Quelle est la probabilité, sachant que la première boîte est occupée, que les particules dans les autres boîtes soient rouges ?

(a) Il y a 2^k répartitions possibles des états, dont 2^k - 1 pour lesquelles au moins une particule est dans l'état 1. La probabilité cherchée est donc 1 / (2^k - 1). En effet, avec les hypothèses faites, les 2^k répartitions sont équiprobables et la probabilité conditionnelle cherchée est le quotient de la probabilité que toutes les particules soient dans l'état 1 par la probabilité qu'au moins une soit dans l'état 1, soit :

$$\frac{1/2^k}{(2^k-1)/2^k} = 1 / (2^k - 1)$$

Si k = 2, on trouve 1/3.

(b) Comme il y a indépendance de l'état d'une particule et de la boîte où elle se trouve, les 2^{k-1} répartitions des k-1 particules qui ne sont pas dans la première boîte sont équiprobables, et la probabilité cherchée est :

$$1 / 2^{k-1}$$

Si k = 2, on trouve 1/2.

On pourrait faire une analogie avec des familles de k enfants, la boîte étant codée par un prénom. Ce peu pertinent : tous les prénoms ne sont pas mixtes. De plus, dans le cas des familles, la liste des prénoms n'est à l'heure actuelle pas fixée… Mais cet exemple des particules, où on donne explicitement le modèle, peut aider à mieux comprendre comment dans ce genre de problème, on peut passer de la probabilité 1/3 à 1/2.

6) Un enfant qui se nomme Dominique. On cherche maintenant la probabilité que cette famille ait deux filles.

Notons :
- D l'événement « un enfant s'appelle Dominique » ;
- DF l'événement « un enfant s'appelle Dominique et c'est une fille » ;
- DG l'événement « un enfant s'appelle Dominique et c'est un garçon ».

On suppose toujours qu'un seul enfant de la famille peut s'appeler Dominique.

Soit x la probabilité, sachant qu'un enfant s'appelle Dominique, que ce soit une fille :

$$x = \text{Prob (DF)/Prob (D)}.$$

D'après la question 4, la probabilité d'avoir deux filles sachant DF est égale à p_4, soit à 1/2.

D'où :

p_6 = Prob (Deuxfilles/D)

$$= \frac{\text{Prob(2filles et DF)} + \text{Prob(2filles et DG)}}{\text{Prob(D)}} = \frac{\text{Prob(2filles / DF)Prob(DF)} + 0}{\text{Prob(D)}}$$

La probabilité cherchée est donc x/2 et vaut 1/4 pour x = 1/2.

Conclusion

Une morale de cette histoire de famille pourrait être que dans une situation de modélisation, et c'est le cas ici, il convient d'analyser avec soin l'information dont on dispose, quitte à envisager plusieurs modèles.

Le contexte choisi l'est pour que les résultats, notamment entre les questions 1 et 4, apparaissent particulièrement paradoxaux : apprécier ou non ce type de paradoxe est une question de goût.

Une autre morale est qu'en pédagogie, s'appuyer sur des exemples ayant trait à la vie courante n'est pas systématiquement heureux, ou favorable à la compréhension du domaine enseigné.