

Enfin en finir avec les problèmes :
**Une mère a deux enfants dont l'un est un
garçon. Quelle est la probabilité que l'autre soit
une fille ?**
et
Le paradoxe des 3 portes

Note Technique
Pierre Dangauthier
Projet EMotion
18 octobre 2004

1 Introduction

Ces deux problèmes sont souvent source d'erreur. Cette note détaille leur solution. Commençons par le problème des enfants, puis terminons par celui des trois portes dans la section 7.

Le premier problème est : *Une mère a deux enfants dont l'un est un garçon. Quelle est la probabilité que l'autre soit une fille ?* Le monde se divise alors en trois, les partisans du $\frac{1}{2}$, ceux du $\frac{2}{3}$ et les éternels indécis prônant l'impossibilité de conclure, l'incomplétude de l'ennocé, la liberté d'opinion, le cinquième amendement, ou le foie de veau à l'ail.

Cette note a pour but :

- de montrer que le problème est bien posé,
- de démontrer que $\frac{2}{3}$ est la seule et unique bonne réponse,
- d'essayer de comprendre pourquoi l'autre réponse est fausse,
- d'exhiber un exemple simple de problème probabiliste pour lequel on se trompe facilement. Les probabilités sont parfois contre-intuitives.

Pour cela, nous expliciterons les 2 raisonnements qui conduisent aux deux réponses, puis nous les formaliserons pour prouver le premier et trouver la faille du second.

2 Ce problème est bien posé

Une mère a deux enfants dont l'un est un garçon. Quelle est la probabilité que l'autre soit une fille ?

Les hypothèses tacites de ce problème sont :

- $P(\text{"avoir un garçon"}) = P(\text{"avoir une fille"}) = \frac{1}{2}$,
- Le sexe du second enfant est indépendant du sexe du premier.
- L'expression "dont l'un est un garçon" signifie bien sûr "dont l'un *au moins* est un garçon".

Ces hypothèses sont sous entendues dans l'ennocé et elles correspondent assez bien à l'expérience réelle.

Nous affirmons d'autre part que ce problème est bien posé car la validation expérimentale est aisée et l'expérience est entièrement définie par l'énoncé. Il suffit de générer une suite de N_1 tirages d'un couple de variables aléatoires bivaluée et de loi uniforme. On éliminera les cas pour lesquels il n'y a aucun garçon (il restera N_2 familles), et on comptera le nombre de familles n_f avec une fille. On verra que le ration $\frac{n_f}{N_2}$ tendra très rapidement vers $\frac{2}{3}$ (loi des grands nombres).

3 Et en vrai

On pourra aussi prendre les statistiques françaises et de compter. Ça marcherait aussi car les hypothèses d'équiprobabilité et d'indépendance sont réalistes.

4 Les raisonnements intuitifs

4.1 Le vrai

Il y a 4 fratries possibles : gg - gf - fg mais pas ff Donc 2 chances sur 3.

4.2 Le faux

Ok, un des 2 est un garçons, mais ça n'influence pas le sexe de l'autre. Donc si je considère l'autre, les chances sont de $\frac{1}{2}$.

4.3 Le problème

De façon très générale, si un problème est bien posé, deux réponses contradictoires ne peuvent pas être juste en même temps. Vu que ce problème est bien posé, un des 2 raisonnements est forcément faux, et c'est "Le faux" qui est faux. Intuitivement il paraît bon, mais on apprend vite qu'en probabilité il vaut mieux se méfier des intuitions (cf problème des portes section 7). Ce raisonnement est intuitif mais il n'est pas rigoureux.

On entend souvent dire que ces 2 raisonnements diffèrent par la prise en compte ou non de l'ordre des enfants. Le "vrai" différencie fg et gf alors que le "faux" se moque de l'ordre. C'est un peu le cas, mais nous verrons que, même en ne considérant pas l'ordre, un raisonnement correct mène bien à $\frac{2}{3}$.

5 Avec rigueur

5.1 Le juste

En probabilités, être rigoureux signifie formaliser le problème (quel est l'espace des possibilités?), énumérer tous les cas possibles en leur donnant leur probabilité. La probabilité solution devient la somme des probabilités des cas favorables. Si tous les cas sont équiprobables, cela revient à calculer le rapport du nombre de cas favorables sur le nombre de cas total.

- Ici l'espace des possibles est l'ensemble des couples de sexes, c'est-à-dire l'ensemble des fratries possibles soit gf gg fg ff. C'est un ensemble de 4 éléments. Ces 4 éléments sont équiprobables : on a 1 chance sur 4 d'avoir fg par exemple. Cette équiprobabilité est due au fait que les sexes des enfants sont indépendants et de probabilité $\frac{1}{2}$ pour g, et $\frac{1}{2}$ pour f.
- Maintenant **sachant** qu'il y a un garçon, les cas possible deviennent seulement gf gg fg : 3 éléments. De plus ce 3 éléments restent équiprobables¹.
- Et les éventualités dans lesquelles "l'autre est une fille" sont gf fg : 2 éléments
- Donc la probabilité est $\frac{2}{3}$.

Nous avons bien tenu compte du rang car nous avons différencié les cas fg et gf.

¹Ceci est intuitif, mais n'est pas trivial à montrer. Monseigneur F.Colas propose une démonstration en introduisant une variable I supplémentaire restreignant l'ensemble des possibles (I=1 ssi ff est impossible soit $P(ff|I=1)=0$), en calculant, par exemple $P(gf|I=1)=\dots=1/3$

5.2 Le faux

Essayons de voir pourquoi l'autre est faux. Dans l'autre on essaie de ne pas tenir compte du rang des enfants. C'est une méthode qui est tout aussi juste, à condition de ne pas se tromper.

Avec cette méthode je vais essayer de mettre en rigueur le raisonnement faux qui donne $\frac{1}{2}$ et ce même raisonnement mais corrigé et qui donne, oh, magie.... $\frac{2}{3}$:

Raisonnement faux formalisé :

- Espace des possibles : { 2FILLES, 2GARCONS, MIXTE } : 3 éléments, car, vu qu'on ne tient pas compte de l'ordre MIXTE = fg ou gf
- SACHANT qu'il y a un garçon il reste { 2GARCONS, MIXTE } : 2 élément
- Eventualités dans lesquelles "l'autre est une fille" : {MIXTE} 1 élément
- Donc probabilité = $\frac{1}{2}$!!

Mais la faute viens du fait que, dans l'espace des possibles, les 3 événements ne sont pas équiprobables. Si on regarde les familles de France avec 2 enfants, la répartition est :

- MIXTE : 50%
- 2GARCONS, : 25%
- 2FILLES, : 25%

Car le cas MIXTE contient en fait 2 sous cas de proba 25%. C'est là l'erreur.

Raisonnement juste :

- Espace des possibles : { 2FILLES, 2GARCONS, MIXTE } avec les probabilités 25%, 25%, 50%
- **Sachant** qu'il y a un garçon il reste { 2GARCONS, MIXTE } avec MIXTE qui compte 2 fois plus que 2GARCONS. Intuitivement, si cet ensemble "pèse" 3 kilos, alors MIXTE en "pèse" 2 et 2GARCONS "pèse" 1 kilo.
- Eventualités dans lesquelles "l'autre est une fille" : { MIXTE } : 1 élément de "poids 2 sur 3"
- donc probabilité = $\frac{2}{3}$

On peut reprocher l'utilisation la notion intuitive, mais non formelle de "poids". C'est uniquement dans un but pédagogique car la démonstration fait appel à la notion de **probabilité conditionnelle** $P(A|B)$ définie par :

$$P(A|B) = P(A \text{ sachant } B) = \frac{P(A \text{ et } B)}{P(B)}$$

Alors :

- $P(2F) = P(2G) = \frac{1}{4}$ et $P(\text{Mix}) = \frac{1}{2}$
- $P(2G \text{ sachant "il y a un g"}) = \frac{P(2G \text{ et "il y a un g"})}{P(\text{"il y a un g"})} = \frac{P(2G)}{P(\text{"il y a un g"})} = \frac{1/4}{3/4} = \frac{1}{3}$
- $P(2F \text{ sachant "il y a un g"}) = \frac{P(2F \text{ et "il y a un g"})}{P(\text{"il y a un g"})} = \frac{P(\text{impossible})}{P(\text{"il y a un g"})} = \frac{0}{3/4} = 0$
- $P(\text{Mix sachant "il y a un g"}) = \frac{P(\text{Mix et "il y a un g"})}{P(\text{"il y a un g"})} = \frac{P(\text{Mix})}{P(\text{"il y a un g"})} = \frac{1/2}{3/4} = \frac{2}{3}$

Ceci peut se reformuler sous forme de programme bayésien (figure 1) pour lequel une seule variable suffit pour définir une fratrie. Soit F le nombre de filles dans la famille : $F \in \{0, 1, 2\}$. D'après les hypothèses, F suit une loi binomiale.

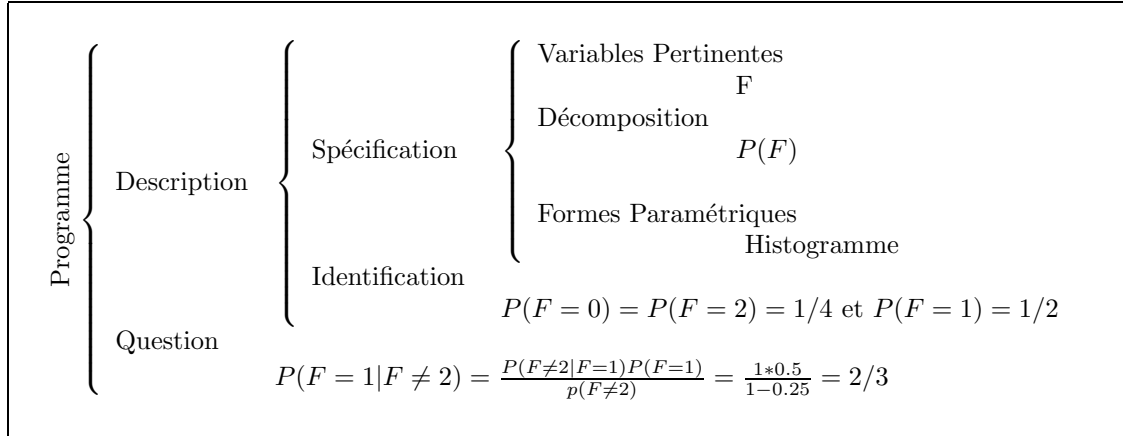


Fig. 1: Le Programme Bayésien donnant la probabilité que le 2ème enfant soit une fille.

6 Conclusion

La seul et unique réponse est $\frac{2}{3}$. L'erreur est dans "un des 2 est un garçons, mais ça n'influence pas le sexe de l'autre".

Car si, justement, ça influence le sexe de l'autre, du moins ça influence la connaissance qu'on a sur le sexe de l'autre. $P(\text{"autre=filles"} \mid \text{sachant "au moins un garçon"})$ n'est pas $\frac{1}{2}$! Car quand j'apprend que "au moins un garçon", je perd le cas "2 filles", et donc je casse la symétrie entre fille et garçon.

Par contre si l'énoncé disait "Une mère a deux enfants dont l'ainé est un garçon" alors la probabilité que le cadet soit une fille serait bien de $1/2$.

7 Un autre problème pas intuitif pour un sous

7.1 Enoncé

Dans un pays où la justice est rendue de façon singulière, un prisonnier a été condamné. Pour déterminer sa peine, il doit choisir entre 3 portes. Derrière deux d'entre elles, c'est l'échafaud, mais derrière la dernière, c'est la liberté ! Le prisonnier désigne la porte n°1. Le geolier, dans sa grande bonté, lui ouvre la porte numéro 3 : c'est l'échafaud !

Il demande : "Voulez-vous changer de porte ?". Vous êtes à la place du prisonnier. Que faites-vous ?

7.2 Réponse fausse

De toute façon, il y avait 1 chance sur 3 que ce soit la porte n°1, et aussi 1 chance sur 3 que ce soit la porte n°2. Donc garder la 1ère ou choisir la 2ème ne change rien à mes chances de survie.

7.3 Oui mais...

Mais si vous faites 1000 fois l'expérience (avec un programme par exemple), vous constaterez qu'en vrai, il vaut mieux changer de porte! En effet le geolier nous importe une information, et il faut en tenir compte pour trouver le meilleur choix.

7.4 Réponse juste, expliquée intuitivement

Si vous ne changez pas de porte car vous pensez que ça ne change rien, vous faites comme beaucoup gens, et vous avez tort! En effet : Avec la stratégie de ne jamais changer de porte, il est clair que vous avez juste une chance sur trois de vous en sortir, puisque il y a une chance sur trois simplement que vous désigniez la bonne porte au départ. Avec la stratégie de toujours changer de porte : si le condamné avait dès le départ choisi la bonne porte, il perd immanquablement. Ce cas se produit avec une probabilité de $\frac{1}{3}$. si le condamné avait choisi une mauvaise porte, ce qui arrive avec une probabilité de $\frac{2}{3}$, le geolier est obligé de montrer la 2ème porte avec l'échafaud. La 3è porte, qui est obligatoirement choisie quand on change, est celle de la liberté. On a donc 2 chances sur 3 de s'en sortir si on change de porte, contre 1 sur 3 si on ne change pas!

Tout l'intérêt des probabilités conditionnelles est dans ce problème. Le gardien apporte une information, dont il faut tenir compte.

7.5 Réponse juste formelle ²

Pour simplifier nous résolvons le problème dans le cas présenté dans l'ennocé. Le prisonnier choisit la première porte et le geolier montre l'échafaud derrière la porte 3. Ce choix ne nous fait pas perdre en généralité et simplifie les notations.

En probabilité, pour ne pas se tromper dans un exercice ou l'intuition est mise en défaut, la méthode est de tout bien écrire. Notons H_i l'hypothèse "la liberté est derrière la porte numéro i ". Le but du prisonnier étant de trouver le H_i dont la probabilité *a posteriori* est maximale. L'énoncé sous-entend les probabilités *a priori* suivantes :

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}.$$

Notons O le numéro de la porte ouverte par le geolier. Si la bonne porte était la 1ère, le geolier peut choisir entre les deux autres, sinon il n'a pas le choix. Ceci s'écrit :

- $P(O = 2|H_1) = \frac{1}{2}$
- $P(O = 3|H_1) = \frac{1}{2}$
- $P(O = 2|H_2) = 0$
- $P(O = 3|H_2) = 1$
- $P(O = 2|H_3) = 1$
- $P(O = 3|H_3) = 0$

Comme on a $O = 3$, il suffit de maximiser sur i :

$$P(H_i|O = 3) = \frac{P(O = 3|H_i)P(H_i)}{P(O = 3)}$$

²Issu du livre de MacKay [Mac00].

Ce n'est même pas la peine de faire le calcul complet car comme $P(H_i)$ et $P(O = 3)$ ne dépendent pas de i , il suffit de trouver le i maximisant $P(O = 3|H_i)$. Les valeurs sont respectivement $\frac{1}{2}, 1, 0$ pour $i = 1, 2, 3$. Donc il H_2 est deux fois plus probable que H_1 . Il faut changer de porte.

Si on veut faire le calcul quand même, on peut calculer $P(O = 3)$ que l'on peut appeler constante de normalisation.

$$P(O = 3) = \sum_i P(O = 3|H_i)P(H_i) = \frac{1/2 + 1 + 0}{3} = \frac{1}{2}$$

Ce qui donne :

$$P(H_1|O = 3) = \frac{1}{3} ; P(H_2|O = 3) = \frac{2}{3} ; P(H_3|O = 3) = 0$$

Références

- [Mac00] D.J.C MacKay. *Information Theory, Inference and Learning Algorithms*. On the WWW, [http ://www.cs.toronto.edu/~mackay/itprnn/book.html](http://www.cs.toronto.edu/~mackay/itprnn/book.html), 2000.