DES PROBABILITÉS POUR DÉCIDER

San Francisco, 22 décembre 1972, une infirmière, nommée Diana Sylvester, est trouvée morte poignardée dans son appartement. La police relève des traces de sperme, mais à cette époque l'analyse génétique n'existait pas. L'enquête de voisinage se dirigea sur un dénommé Robert Baker qui semblait le suspect idéal, mais l'affaire fut classée faute de preuves et le dossier dormit trente ans aux archives des affaires non résolues.

En 2003, la police de San Francisco reçut une subvention destinée à l'usage de nouvelles technologies. L'idée était de rouvrir les anciens dossiers contenant un prélèvement d'ADN et de confronter cet ADN à une grande base de données qui venait d'être constituée et qui rassemblait les profils génétiques de criminels californiens.

On rouvrit le dossier de Diana qui contenait une lame avec un peu de sperme prélevé sur le corps de Diana. Le sperme en question était endommagé et ne faisait apparaitre que partiellement le code génétique de l'assassin. Pour les cinq *locus* identifiables, la RMP (*Random match probability*) est estimée à 1 pour 1,1 millions. La probabilité qu'un individu possède les mêmes caractéristiques que les caractéristiques lisibles de l'assassin est estimée à 1/1.100.000.

La base de données des délinquants californiens comportait 338.000 profils et on trouva un individu, du nom de John Puckett, dont l'ADN avait les cinq caractères de l'ADN de l'assassin.

Joli procès en perspective, on cherchait un individu ayant le même ADN que l'assassin, on en trouve un, et c'est quand même plutôt rare d'avoir cet ADN : une chance sur un million. Qu'est-ce que ça nous dit sur la probabilité de la culpabilité de Puckett ? Il n'y a qu'une personne sur un million qui a cet ADN. Je répète comme un procureur porrait le faiire au prétoire.

Le biais du procureur

C'est la lecture du livre de Vincent Berthet[1] qui m'a fait rouvrir le livre *Les Maths au Tribunal* de Leila Schneps et Coralie Colmez où le cas de John Puckett est développé[2]. Vincent Berthet étudie divers processus de raisonnements irrationnels, les biais de raisonnement. Il cite et reprend les travaux de Kahneman et Tversky[3] et il y ajoute le *biais du procureur*. C'est une inversion des probabilités conditionnelles : au lieu de se demander quelle est la probabilité que l'accusé soit innocent, compte tenu de son dossier, on se demande quelle est la probabilité qu'il ait ce dossier s'il est innocent. Dans le cas du crime de San-Francisco, le dossier c'est l'ADN. Et on a une « probabilité de procureur » : la

---

[1] BERTHET (Vincent), *L'erreur est humaine,* CNRS Éditions, 2018

[2] SCHNEPS (Leila) & COLMEZ (Coralie), *Les Maths au Tribunal,* Seuil, 2015, titre original : *Math on Trial. How Numbers get Used and Abused in the Courtroom,* Basic books, 2013

[3] KAHNEMAN (Daniel), *Système 1, système 2*, Flammarion, 2016.

probabilité d'un sur un million qu'un individu quelconque, innocent ou coupable, ait l'ADN en question.

L'analyse de l'ADN de John Puckett, c'est analogue à un test médical : avez-vous l'ADN de l'assassin ? Il peut y avoir des faux positifs. C'est ce qui m'a donné l'envie d'écrire ce texte. La période s'y prête, on parle beaucoup de tests, mais on parle peu des faux positifs ou négatifs.

Pour tenir le lecteur en haleine, je précise tout de suite qu'il n'y avait pas d'impossibilité technique à ce que Puckett fût l'assassin. Âgé de soixante-dix ans en 2003, il avait la quarantaine et habitait la région au moment du meurtre de Diana.

Tests biologiques

Un test biologique donne un résultat positif P ou un résultat négatif N. Les médecins attribuent deux caractéristiques à un test biologique :

- la **sensibilité** qui est la probabilité P(P|M) (on dit « P sachant que M ») que le test soit positif pour un malade,
- la **spécificité** qui est la probabilité P(N|S) que le test soit négatif sur un sujet sain.

Au lieu de « probabilité », on peut parler de proportions. La sensibilité nous dit, sur l'ensemble des malades testés, combien de malades ont un test positif. Le complémentaire 1-P(P|M) est la proportion P(N|M) de tests négatifs sur les malades testés, ce sont les **faux négatifs**.

La spécificité, c'est le paramètre symétrique : quelle proportion de personnes saines testées ont un test négatif. Le complémentaire 1-P(N|S) = P(P|S) est la proportion de **faux positifs**.

La proportion P(P|M), la sensibilité, c'est la probabilité du procureur. Prenons un test dont la sensibilité est 90% (c'est un minimum). Si le test est positif, ai-je 90 chances sur 100 d'être contaminé ? Pas du tout Monsieur le procureur. Ce qui est significatif, c'est la proportion P(M|P) des malades parmi les personnes testées positives.

Heureusement, les mathématiciens savent passer de l'un à l'autre en utilisant la remarque d'un pasteur presbytérien britannique du XVIII-ème siècle, Thomas BAYES (Londres 1702 - Tunbridge Weels 1761). Précisons maintenant que nous raisonnons sur une population de personnes testées, les personnes non testées n'interviennent pas.

Bayes remarque qu'il y a deux façons de calculer le nombre N(M et P) des malades testés positifs ou, ce qui revient au même, des testés positifs qui sont malades :

• le nombre N(M) de malades multiplié par la proportion P(P|M) de positifs parmi les malades

$$N(M \text{ et } P) = P(P|M) \times N(M) ,$$

• ou bien, le nombre N(P) de positifs multiplié par la proportion P(M|P) de malades parmi les positifs

$$N(M \text{ et } P) = P(M|P) \times N(P).$$

D'où l'égalité

$$P(P|M) \times N(M) = P(M|P) \times N(P),$$

connue sous le nom de « formule de Bayes » ou de « probabilité des causes ».

On peut ainsi exprimer la proportion qui nous intéresse

$$P(M|P) = P(P|M) \times N(M) / N(P).$$

• $P(P|M)$ c'est la sensibilité,
• $N(M)$: on a besoin de $N(M)$, le **nombre de malades dans la population testée[4]**, pour pouvoir évaluer la significativité d'un test positif. C'est la règle fondamentale que nous serinons à nos étudiants.
• $N(P)$: le nombre de tests positifs, on le constate tout simplement. Mais on peut aussi essayer de le prévoir en additionnant les vrais positifs et les faux positifs :

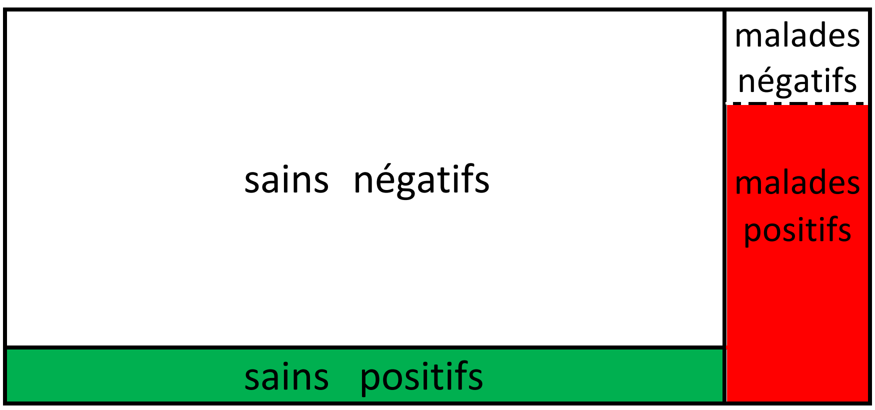
$$N(P)= P(P|M) \times N(M) + P(P|S) \times N(S).$$

Voici un petit calcul pour un test de sensibilité 90%, de spécificité 97%, et une proportion de 1 malade pour 1000 sujets testés[5]

$$N(P)= P(P|M) \times N(M) + P(P|S) \times N(S).$$
$$= 0{,}9 \times 1 + 0{,}03 \times 999 = 31$$
$$P(M|P) = P(P|M) \times N(M) / N(P) = 0{,}9 \times 1 / 31 = 2{,}9\%$$

Vous voyez le désastre ! Il n'ya pas plus de 3% de malades parmi les tests positifs. Et vous voyez pourquoi il ne faut pas se laisser tromper par les 90% de sensibilité $P(P|M)$. Dans le calcul surligné, les 31 positifs sont la somme d'un malade (0,9 x 1) et de 30 personnes saines (0,03 x 999).

On peut voir cela sur un dessin. Les tests positifs sont en couleur : rouge pour les malades, verts pour les sujets sains. Les faux positifs sont une faible proportion pour les sujets sains, mais ils peuvent être une forte proportion parmi les tests positifs si les sujets sains sont très majoritaires.



[4] Voir en annexe un texte de Henri Poincaré sur cette question.
[5] Le nombre de 66.000 cas recensés en France est atteint (et dépassé) le 3 avril 2020.

En 2020, les tests qui ont été déployés dès le mois de mars avaient une sensibilité parfois médiocre (de 70% à 95%) tandis que les spécificités étaient toujours supérieures à 95%. Il y avait alors très peu de malades du covid.

Pour que les tests soient utiles, il faut bien sûr diminuer les faux positifs (qui sont 97% des positifs dans l'exemple numérique ci-dessus). Pour cela on doit améliorer la spécificité et la sensibilité d'une part, et d'autre part, et surtout, limiter les tests à une population à risque, par exemple aux sujets qui ont des symptômes.

Rassurons-nous par un nouveau calcul avec des tests ayant 99% de sensibilité et de spécificité et une population infectée à 5% comme c'est le cas en France début février 2021.

$$P(M|P) = 0{,}99 \times 0{,}05 / (0{,}99 \times 0{,}05 + 0{,}01 \times 0{,}95) = 0{,}84$$

Une proportion de 84% de malades parmi les tests positifs. C'est pas mal, mais tempérons un peu notre optimisme en pensant que les personnes qui se savent déjà infectées ne vont pas faire un test.

Au printemps dernier, on recommandait un test uniquement aux personnes présentant des symptômes de l'infection. L'objectif était alors d'hospitaliser le plus tôt possible les personnes atteintes sans attendre un développement dramatique de la maladie.

En décembre 2020, la Région Rhône-Alpes-Auvergne a financé un dépistage massif dont l'objectif ni les résultats n'ont été clairement explicités[6]. En janvier 2021, un dépistage massif à Saint-Étienne est piloté par une équipe hospitalière. Là, l'objectif est clairement avoué, l'équipe le dit à la télé, elle veut faire une publication.

L'ADN est-il une preuve ?

Revenons à San Francisco. Nous allons voir que notre analogie avec les tests biologiques ne donne rien. En effet, on a testé une population de 338.000 personnes. Le test P, c'est l'ADN, et le caractère M, c'est d'être coupable.

On voit que P(P|M)=1, le coupable a bien l'ADN, c'est la sensibilité du test. La spécificité, c'est 0,999.999, une chance sur un million d'avoir un faux positif. On essaye d'écrire l'égalité de Bayes

$$P(M|P) = P(P|M) \times N(M) / N(P).$$

Le nombre N(P) de positifs est égal à 1 (c'est John Puckett), mais le nombre N(M), dont on sait qu'il est essentiel, on ne le connait pas. Il y a deux possibilités :

- N(M) = 1, l'assassin est parmi les 338.000, et c'est bien sûr Puckett,
- N(M)=0, l'assassin n'est pas parmi les 338.000 et Plucket est innocent.

On pourrait conclure si l'on était sûr que l'assassin est bien dans la population examinée. Mais c'est justement la question.

Il n'y avait pas besoin de probabilités pour coonclure, mais seulement d'un peu de logique :

[6] La campagne est lancée avant Noël pour « protéger nos proches ». Bilan : 8 millions d'habitants, 626.000 tests et 29.000 tests positifs.

- si l’assassin est dans la base de données, c’est Puckett, le seul à avoir l’ADN,
- si l’assassin n’est pas dans la base de données, ce n’est pas Puckett, bien sûr.

Restons à San Francisco où une avocate a été désignée pour la défense de John Puckett. Bicka Barlow avait travaillé dans la médecine légale et elle était à présent membre du service de la défense publique de la ville de San Francisco. Elle avait depuis longtemps des doutes sur la méthode des coïncidences dans les bases de données et elle connaissait les travaux de Kathrine Trayer qui avait fait une gigantesque étude statistique sur une base de données génétiques de l’état de l’Arizona. La question était de distinguer entre « probabilité d’avoir un ADN donné » et « probabilité que deux individus aient le même ADN ».

Sur une base de données de l’ADN de 65.000 personnes, on examine les coïncidences de 9 locus. Pour 9 locus, le RMP est de 1 sur 13 milliards. Cependant la base de données contient 122 paires d’individus ayant les mêmes 9 locus. Comment est-ce possible ?

Les anniversaires

C’est l’occasion de faire une digression.

C’est un petit jeu que l’on peut pratiquer en classe, par exemple dans une classe de 25 élèves. On est étonné qu’il soit fréquent que deux élèves de la classe aient leur anniversaire le même jour alors qu’il y a 25 élèves dans la classe et 365 jours dans l’année.

1. Quelle est la probabilité que l’un des élèves ait son anniversaire le 15 juin ?
Réponse : la probabilité qu’un élève donné n’ait pas son anniversaire le 15 juin est 364/365.
La probabilité qu’aucun élève de la classe n’ait son anniversaire le 15 juin est égale à

(364/365)^25 = 0,9337.

On a multiplié les probabilités d’évènements indépendants[7].
La probabilité qu’au moins un élève ait son anniversaire le 23 janvier est

1 - 0,9337 = 0,0663

c’est-à-dire 6,63%.

2. Quelle est la probabilité que deux élèves aient leur anniversaire le même jour ?
Réponse : cherchons la probabilité que tous les anniversaires soient différents dans une classe de n élèves. Ce n’est pas une probabilité que l’on calcule, mais la proportion P(n) parmi tous les cas possibles. On met en ordre les élèves et on les regarde un à un.

P(1) = 1, s’il y a un seul élève, il n’y a pas de coïncidence.

P(2) = 364/365, pour le deuxième élève, il y a 364 jours anniversaires possibles qui sont différents du jour anniversaire de l’élève numéro 1,

P(3) = 363 x 364 / 365^2, pour le troisième élève, il y a deux jours interdits,

P(n+1) = P(n) x (365-n) /365, pour le (n+1)-ième élève, il y a n jours interdits.

[7] La notation 2^3 signifie : 2 à la puissance 3.

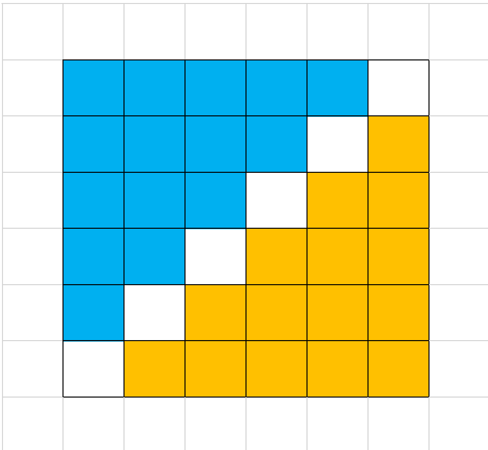
Sur mon tableur, voici les résultats que j'obtiens :

n	1	2	3	4	5	6	7	8	9	10	11	12	13
P(n)	1	0,99	0,99	0,98	0,97	0,96	0,94	0,93	0,91	0,88	0,86	0,83	0,81

n	14	15	16	17	18	19	20	21	22	23	24	25	26
P(n)	0,78	0,75	0,72	0,68	0,65	0,62	0,59	0,56	0,52	0,49	0,46	0,43	0,40

Dès que l'effectif de la classe atteint 23 élèves, la probabilité P(n) descend au-dessous de 50% et il y a plus d'une chance sur deux que deux élèves aient le même jour anniversaire. Dès un effectif de 57 élèves, cette probabilité dépasse 99%.

3. Un calcul erroné. Voici un calcul erroné de cette probabilité des anniversaires. Dans une classe de 25 élèves, nous considérons toutes les « paires » d'élèves. Il y a 25x25 façons de prendre deux élèves au hasard, soit 625. On soustrait les 25 tirages où l'on a pris deux fois le même élève, il reste 600. Et on divise par deux car chaque paire a été tirée deux fois. On peut illustrer cela par un dessin pour une classe de 6 élèves.



On arrive ainsi à 300 paires de 2 élèves. À chacune de ces paires correspond une paire de dates de l'année. Sur ces paires de dates, il y en a une sur 365 qui est double, soit une probabilité p = 1/365. La probabilité qu'une paire de dates ne soit pas double est égale à

$$q = 1 - p = 364/365 = 0{,}99726.$$

La probabilité que les 300 paires d'élèves évitent les doubles dates serait

$$Q(25) = q^{300} = 0{,}43909,$$

alors que le raisonnement précédent (correct) donnait 0,4313.

On peut se poser deux questions auxquelles je vais répondre :

- pourquoi ces deux calculs donnent-ils des résultats voisins ?
- pourquoi le second raisonnement est-il erroné ?

Commençons par la deuxième question, et limitons-nous à trois élèves A, B et C. Le fait que chacun des trois couples ait des dates d'anniversaires différentes ne sont pas des évènements indépendants et on ne peut pas se contenter d'élever au cube la probabilité *q*.

Supposons que les élèves A et B aient des anniversaires différents. Pour l'élève C, la probabilité que son anniversaire soit différent de celui de l'élève A est toujours égale à q. Mais la probabilité que l'anniversaire de C soit différent des deux anniversaires (différents) de A et de B est égale à 363/365 et non à q^2 = (364/365)^2.

On reprends ces deux calculs pour répondre à la première question. On compare P(3) et Q(3) .

$$Q(3) = \left(1 - \frac{1}{365}\right)^3 \;;\quad P(3) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right).$$

On apprend en classe terminale que, lorsque le nombre x est très petit, l'expression (1 - 2 x) est très voisine de $(1-x)^2$.

Alors, pourquoi faire un calcul erroné ? Avant l'année 1970, nous n'avions pas de calculette ni d'ordinateur. Pour faire le calcul de P(25), il fallait faire 23 multiplications, par 363, 362, …, et 24 divisions, toutes par 365. Au lieu de faire ce calcul, j'ai préféré recourir au calcul de l'approximation Q(25) que l'on fait facilement avec une table de loogarithmes.

Retournons en Arizona : une base de données de 65.000 personnes. Pour le raisonnement correct, il faudrait faire 65.000 itérations et, comme je ne suis pas spécialiste, je ne sais pas faire avec mon petit ordinateur de bureau. Le calcul de Q(n), avec n=65.000, est plus approprié.

nombre de couples n x (n-1)/2 soit plus de 2 milliards,
multiplier par 715, le nombre de choix de 9 locus parmi 13,
on est à plus de 1.500 milliards de possibilités de coïncidence
pour 9 locus, le RMP est égal à 13 milliards : chaque couple a une chance sur 13 milliards de présenter une coïncidence,
Je ne fais pas le calcul, mais il n'est pas étonnant de rencontrer 120 paires d'ADN coïncidant

Le nombre de couples augmente comme le carré du nombre d'individus. Pour 1 million de personnes, cela donnerait 500 milliards de couples.

Retour à San-Francisco

Après cette instructive digression, revenons à San Francisco, l'avocate essaya de parler de coïncidences pour faire réviser le chiffre de 1 sur 1 million. Elle avait remarqué, comme nous venons de le voir avec l'exemple des anniversaires ou de la base de l'Arizona, que la probabilité que deux personnes aient le même ADN partiel (les cinq locus) est beaucoup plus élevée que 1 sur 1 million. Mais le procureur la remit sur le droit chemin : la question est de savoir si un individu a un ADN donné, celui de l'assassin, et c'est bien une chance sur 1 million. Ce n'est pas le fait que deux individus aient un même ADN qui est en question.

La base de données

L'accusation interprète le rapport de 1 sur 1 million comme la probabilité que Puckett **ne soit pas** celui dont provient l'échantillon. Ça ne lui fait pas un bon dossier : 999.999 chances sur un million qu'il soit coupable.

L'avocate Bicka Barlow, qui n'aime pas les bases de données, entreprend une autre démarche. L'ADN de Puckett figurait dans une base de données de 338.000 individus. Notre proportion 1 sur 1 million fait qu'il y a une chance sur 3 de trouver le bon ADN dans une base de 338.000. C'est ainsi que Puckett a été trouvé. Selon Bicka Barlow, cette extraction de Puckett de la base de données n'a rien à voir avec le meurtre de Diana ; il y avait statistiquement 1 chance sur 3 que l'on trouve quelqu'un avec l'ADN de l'assassin dans une base de 338.000 individus, comme dans n'importe quelle base de données de 338.000 individus.

Avant le début du procès, le juge décida qu'il serait trop déroutant de présenter devant la cour les différents arguments mathématiques qui, d'ailleurs, semblaient se contredire. Le jury a condamné Puckett à la prison à vie. Il faut dire que Puckett n'était pas par hasard dans la base de donnée de criminels de Californie : il avait à son actif plusieurs viols, commis dans des conditions voisines du viol (ou plutôt de la fellation contrainte) de Diana Sylvester. C'est une condamnation à la française sur des présomptions et, en plus, trente ans après les faits supposés.

Réfléchissons

Dans leur livre *Les Maths au tribunal*, Leila Schneps et Coralie Colmez, essayent un autre raisonnement, qui est développé dans Wikipédia à l'article « empreinte génétique ». Elles délimitent la population de toutes les personnes susceptibles d'être l'assassin pour appliquer dans cette population le 1 sur 1 million de la trace d'ADN. Cette population est constituée d'hommes, californiens, agés de plus de 65 ans et délinquants sexuels. Elles dénombrent 18.750 individus ayant ces caractères. Dans cette population, il y a une chance sur un million qu'un individu ait la trace d'ADN de l'assassin.

$$18.750/1.000.000 = 1/70$$

Il y a une chance sur 70 qu'un californien, âgé de plus de 65 ans et récidiviste, ait le « dossier », le test positif.

Sans le dire explicitement (respect de la chose jugée ?) les auteures suggèrent que cela signifie qu'il y a une chance sur 70 que l'assassin soit quelqu'un d'autre que Puckett. En prolongeant leur non dit, cela voudrait signifier qu'il y a 69 chances sur 70 que Puckett soit l'assassin.

Mais est-ce une raison suffisante pour condamner Puckett ?

Et après

Les tests ADN sont devenus une routine pour les enquêteurs comme les empreintes digitales l'étaient devenues il y a un siècle. Le profil ADN, s'il est complet et non partiel comme à San Francisco en 1972, permet une identification certaine de l'individu.

Pour le « grand public » comme pour beaucoup de journalistes et peut-être aussi de procureurs, l'ADN est une preuve, mais ce n'est heureusement pas toujours une preuve pour les juges qui instruisent à charge et à décharge.

Voici ce que j'ai vu et entendu par hasard à la télévision. Le 7 février sur la chaine RMC, Frédérique Lantieri interroge la juge d'instruction en charge du dossier des « souris vertes ». La juge explique que la seule trace laissée par les gangsters est une cagoule sur laquelle on trouve un peu d'ADN mitochondrial[8]. On compare cette trace d'ADN à l'ADN de la mère d'un suspect et c'est positif. « C'est la preuve ! » dit Frédérique Lantieri, « C'est un élément » rectifie la juge.

Ajoutons que l'ADN permet à coup sûr de disculper.

BIBLIOGRAPHIE

SCHNEPS (Leila) & COLMEZ (Coralie), *Les Maths au Tribunal,* Seuil, 2015, titre original : *Math on Trial. How Numbers get Used and Abused in the Courtroom,* Basic books, 2013

BERTHET (Vincent), *L'erreur est humaine,* CNRS Éditions, 2018

KAHNEMAN (Daniel), *Système 1, système 2*, Flammarion, 2016.

ANNEXE

Probabilités des effets ou des causes

Henri POINCARE, *La science et l'hypothèse*, Flammarion, Paris, 1943

****

p.221 : *Il arrive souvent qu'au lieu de chercher à deviner un événement d'après une connaissance plus ou moins importante de la loi, on connaisse les événements et qu'on cherche à deviner la loi ; qu'au lieu de déduire les effets des causes, on veuille déduire les causes des effets*. […]

Je joue à l'écarté avec un monsieur que je sais parfaitement honnête ; il va donner ; quelle est la probabilité pour qu'il tourne le roi ? c'est 1/8 ; c'est là un problème de probabilité des effets. Je joue avec un monsieur que je ne connais pas ; il a donné 10 fois et il a tourné 8 fois le roi ; quelle est la probabilité pour que ce soit un grec [un tricheur] ? C'est là un problème de probabilité des causes.

p.237 : *Pour mieux m'expliquer, je reviens à l'exemple du jeu d'écarté, cité plus haut ; mon adversaire donne pour la première fois et il tourne le roi ; quelle est la probabilité pour que ce soit un grec ? Les formules ordinairement enseignées donnent 8/9, résultat évidemment bien surprenant. Si on les examine de plus près, on voit qu'on fait le calcul comme si, avant de nous asseoir à la table de jeu, j'avais considéré qu'il y avait une chance sur deux pour que l'adversaire ne fût pas honnête. Hypothèse absurde, puisque, dans ce cas, je n'aurais certainement pas joué avec lui ; et c'est ce qui explique l'absurdité de la conclusion*.

****

Comment trouve-t-on cette probabilité égale à 8/9 ?

[8] L'ADN mitochondrial est beaucoup moins discriminant que l'ADN nucléaire.

On imagine un espace probabilisé à 16 cases également probables, les 8 cases du grec et les 8 cases de l'honnête joueur. Les 8 cases du grec retournent un roi, mais une seule case du joueur honnête est un roi. Ainsi, sur 9 rois retournés, 8 sont le fait du grec et 1 le fait du joueur honnête.

Avec des calculs, on cherche la probabilité P(G|R) (probabilité de grec sachant que roi).

$$P(G|R)\ P(R) = P(G \cap R) = P(R|G)\ P(G)$$

On a P(R|G) = 1 ; Poincaré nous dit que l'hypothèse implicite est P(G) = 1/2. Cette donnée intervient dans la relation ci-dessus, mais elle intervient aussi dans le calcul de P(R) qui n'est pas 1/8 à cause de la possibilité de triche.

$$P(R) = P(R|H)\ P(H) + P(R|G)\ P(G)$$
$$P(R) = (1/8)\ (1/2) + 1\ (1/2) = 9/16$$

$$P(G|R) \times 9/16 = 1/2$$
$$P(G|R) = 8/9$$

Pour lever l'hypothèse implicite, notons p la probabilité *a priori* que le joueur soit un tricheur.

$$P(G\,|\,R) = \frac{p}{(1-p)/8 + p} = \frac{8p}{1+7p}$$

Ci-dessous la courbe représentative de la fonction y = 8x/(1+7x) :