LOGIQUE & CALCUL

L'embarrassant paradoxe de Simpson

Les statistiques conduisent à des décisions fondées et rationnelles… à moins qu'un paradoxe ne s'en mêle, comme cela arrive avec certains jeux de données.

Jean-Paul DELAHAYE

Un bon paradoxe est un paradoxe dont on ne réussit jamais à se débarrasser. Quand vous croyez en avoir trouvé la clef, une remarque vous fait découvrir que rien n'est résolu. Les paradoxes de Zénon à propos de l'impossibilité du mouvement sont de tels paradoxes. Mais le plus élémentaire de tous est le paradoxe de Simpson dont on imagine des solutions… qui conduisent à d'autres paradoxes !

Sans cesse, des scientifiques et des utilisateurs de statistiques tombent dans les pièges qu'il tend. Chaque année, paraissent des articles qui tentent de déterminer son sens profond et la façon dont on doit le traiter. Malgré cette littérature abondante, il n'est pas certain que l'on détienne une solution entièrement satisfaisante pour se libérer de cette récalcitrante absurdité.

Imaginons la situation suivante. On mène des tests en double aveugle sur un nouveau médicament traitant la maladie grave MG. On a traité 160 patients, dont 80 ont reçu le médicament, et les 80 autres un placebo.

Le taux de guérison varie selon que l'on considère les malades ayant pris le médicament ou ceux ayant pris le placebo (*voir le tableau 1 ci-contre*). Parmi les 80 patients ayant pris le médicament, 40 ont été guéris (50 %). Parmi les 80 patients ayant reçu le placebo, seuls 32 ont été guéris (40 %). Ces résultats suggèrent que le médicament est efficace. Il faut donc le prescrire pour soigner les patients atteints de la maladie MG.

Mais en analysant plus en détail les données et en considérant le sexe des personnes ayant participé aux tests, on a une surprise : parmi les hommes, le placebo réussit mieux que le médicament, et il en va de même parmi les femmes. La somme des deux tableaux 2 et 3 redonne bien le tableau 1.

1

Total	Guéri	Non guéri	Taux de guérison
Médicament	40	40	50 %
Placebo	32	48	40 %

2

Hommes	Guéri	Non guéri	Taux de guérison
Médicament	36	24	60 %
Placebo	14	6	70 %

3

Femmes	Guéri	Non guéri	Taux de guérison
Médicament	4	16	20 %
Placebo	18	42	30 %

Les trois tableaux sont compatibles, aucune erreur ne s'est produite. Les résultats sont sans appel : chez les hommes, le placebo est meilleur que le médicament ; chez les femmes, le placebo est meilleur que le médicament. Pourtant, en regroupant hommes et femmes, le médicament produit de meilleurs résultats que le placebo. On se trouve dans un cas du paradoxe de Simpson : la fusion de données concluant individuellement dans un sens – l'inutilité du médicament testé – donne des résultats concluant dans le sens inverse, l'utilité du médicament.

La plus lointaine mention d'un cas analogue remonte à 1899, où le mathématicien anglais Karl Pearson décrivit des données équivalentes. Plus tard, en 1903, Undy Yule redécouvrit le phénomène, et le Britannique Edward Simpson (né en 1922) écrivit en 1951 un article où cette singularité statistique était soigneusement étudiée et discutée. Assez injustement aujourd'hui, ce type de données paradoxales porte son nom.

De nombreux cas réels présentent cette inversion de résultat lorsqu'on regroupe plusieurs catégories complémentaires en une seule. Chaque année, on découvre de nouveaux exemples produisant étonnement et incrédulité. On a rencontré le paradoxe à propos des taux d'admission des filles dans divers départements de l'Université de Berkeley en 1973 : elles étaient meilleures que les garçons dans la plupart des départements de l'université, mais quand on fusionnait les résultats, leur taux général d'admission à l'université était inférieur à celui des garçons. De nombreux cas en médecine ont été rapportés. Le paradoxe a aussi été rencontré à propos de l'analyse des matchs de basket-ball, en démographie, dans l'étude des risques d'accidents, etc. (*voir l'encadré 1*).

Explication arithmétique

Avant de tenter de savoir s'il faut oui ou non prescrire le médicament testé dans le cas des tableaux 1-3, étudions l'arithmétique du paradoxe. Si l'on note A, B, C et D les quatre nombres du premier tableau, a, b, c et d les quatre nombres du second tableau, a', b', c' et d' les quatre nombres du troisième tableau, on a les relations suivantes :

$$A = a + a' ; B = b + b' ; C = c + c' ; D = d + d' ;$$
$$A/B > C/D ; a/b < c/d ; a'/b' < c'/d'.$$

L'étonnement vient de ce que l'on croit que [$a/b < c/d$ et $a'/b' < c'/d'$] entraîne $A/B < C/D$, ou encore avec seulement les petites lettres : [$a/b < c/d$ et $a'/b' < c'/d'$] entraîne $(a + a')/(b + b') < (c + c')/(d + d')$.

Or aucune démonstration n'établira cette implication entre inégalités arithmétiques, puisque justement les nombres des tableaux 1-3 vérifient les deux premières inégalités et pas la troisième : $36/24 < 14/6$ et $4/16 < 18/42$, mais $40/40 > 32/48$ (soit $1,5 < 2,33$ et $0,25 < 0,42$, mais $1 > 0,66$).

Sur le plan arithmétique, il n'y a pas le moindre paradoxe – seulement une légère surprise dont on se libère en analysant l'exemple qui nous montre que même si $a/b < c/d$ et $a'/b' < c'/d'$, il ne faut pas en conclure que $(a+a')/(b+b') < (c+c')/(d+d')$.

Une représentation graphique des données *(voir l'encadré 2)* illustre pourquoi il ne faut pas s'étonner de rencontrer des jeux de huit paramètres produisant le « paradoxe ». Cette représentation conduit d'ailleurs à découvrir des cas encore plus simples de jeux de données présentant la prétendue anomalie. En imposant aux variables d'être des entiers entre 1 et 4, on trouve pour les huit paramètres quatre solutions qui sont : (1, 2, 3, 4, 4, 2, 3, 1) (1, 3, 2, 4, 4, 3, 2, 1) (4, 2, 3, 1, 1, 2, 3, 4) (4, 3, 2, 1, 1, 3, 2, 4).

Lorsqu'on impose aux variables d'être des entiers entre 1 et 5, le nombre de solutions explose : il y en a 232. Pour l'intervalle de 1 à 6, il y en a 1 370. Pour les intervalles suivants, le nombre de solutions est 8 126, puis 28 252, puis 86 140, etc. La proportion de jeux de huit entiers pris entre 1 et n vérifiant les inégalités du paradoxe (ou les inégalités inverses qui sont aussi paradoxales) tend vers 0,9606 % quand n tend vers l'infini.

Comment raisonner ?

Tout n'est cependant pas réglé pour autant et, dans la réalité, un médecin face aux données des trois tableaux indiqués doit prendre une décision : oui ou non, faut-il prescrire le médicament qui semble efficace (d'après le tableau 1) et qui semble moins bon que le placebo (d'après les tableaux 2 et 3) ?

Imaginez-vous à la place du médecin. Comment allez-vous raisonner ? Plusieurs attitudes sont possibles.

1. Trois exemples réels

Pris parmi des dizaines d'autres, voici trois cas réels où le paradoxe de Simpson a produit des surprises et a contraint à analyser plus finement ce que des manipulations statistiques imprudentes faisaient apparaître.

Aptitudes en progrès pour tous, mais dont la moyenne est stable.
Aux États-Unis, au cours des années 1981-2002, la moyenne des notes obtenues par les élèves soumis au test linguistique SAT *(Scholastic Aptitude Test)* a été stable, avec un résultat toujours proche de 504 points. Pourtant, durant cette période, la moyenne pour le test mesuré séparément par groupe ethnique s'est accrue dans chaque groupe d'au moins huit points.

Gerald Bracey, de l'Université George Mason, a étudié cet étrange phénomène et a conclu qu'on était là en présence d'un cas du paradoxe de Simpson.

La clef de l'explication se trouve dans le fait que pendant la période considérée, le nombre d'élèves blancs soumis au test a proportionnellement décru, et que la proportion d'élèves appartenant à des minorités (réussissant en général moins bien le test linguistique) a augmenté, conduisant au total à une stagnation des résultats globaux qui, pourtant, progressaient dans chaque catégorie.

Mauvais taux de mortalité dans chaque tranche d'âge, mais globalement bons.
En 1986, Joel Cohen, de l'Université Rockefeller à New York, a identifié un cas réel de paradoxe de Simpson dans le domaine de la démographie. Il s'est intéressé à la mortalité au Costa Rica et en Suède (connue pour son excellente espérance de vie). Sans surprise, il a trouvé qu'en 1960, le taux de mortalité des femmes dans toute tranche d'âge fixée était supérieur au Costa Rica, comparé au taux équivalent en Suède. Pourtant, le taux de mortalité général des femmes au Costa Rica était inférieur à celui de la Suède.

L'explication est encore liée au paradoxe de Simpson, et elle provient de la structure différente des populations. La population du Costa Rica est beaucoup plus jeune en moyenne que celle de la Suède, et donc les jeunes classes d'âge (qui ont un faible taux de mortalité) pèsent plus dans la moyenne pour le Costa Rica que pour la Suède, conduisant à un taux de mortalité global assez faible au Costa Rica, malgré un taux assez mauvais dans chaque tranche d'âge.

En roulant plus vite, vous aurez moins d'accidents.
Gary Davis, de l'Université du Minnesota, étudiait en 2004 la relation entre le nombre d'accidents piéton-véhicule et la vitesse moyenne des véhicules en divers endroits d'une ville.

Il proposa une modélisation de cette relation qu'il voulait utiliser pour évaluer l'intérêt de limiter plus sévèrement la vitesse autorisée des véhicules. De façon inattendue, le modèle montrait qu'en faisant passer la limite de 30 miles par heure à 25, le nombre d'accidents augmenterait.

Là encore, bien que pour chaque type de sites, le nombre d'accidents diminuait quand la vitesse limite était réduite, une agrégation malheureuse des données (qui ne prenait pas en compte que le nombre d'accidents était bien plus rare dans les zones résidentielles) engendrait une conclusion absurde.


Étudiants en Virginie

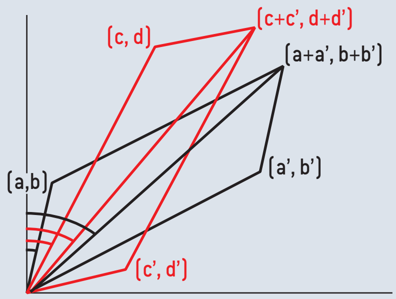

Rue au Costa Rica

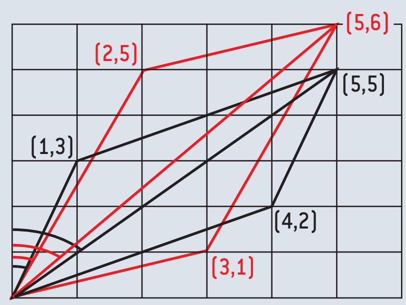

Rue en Suède

Limitation de vitesse

© Michael Rosskothen/shutterstock.com

2. Interprétation graphique





Le paradoxe de Simpson apparaît lorsque $a/b < c/d$, $a'/b' < c'/d'$ et que néanmoins $(a + a')/(b + b') > (c + c')/(d + d')$. Les graphiques où les rapports a/b sont une fonction monotone des angles indiqués en fournissent la preuve graphique. Grâce à cette représentation, on trouve un exemple qui prouve que les trois inégalités, même dans des cas simples impliquant de petits nombres, peuvent être vraies simultanément.

Point de vue 1. Quand j'ignore si c'est un homme ou une femme (car par exemple je suis en train de traiter le cas d'un malade anonyme), je ne tiens compte que de la statistique générale qui est celle qui s'applique dans ce cas. Je donne donc le médicament, car je sais qu'il conduit à la guérison dans 50 % des cas d'après le tableau 1, alors que le placebo ne conduit à la guérison que dans 40 % des cas. En revanche, lorsque j'ai plus de précisions sur la personne à traiter et que je sais s'il s'agit d'un homme ou d'une femme, je regarde la statistique correspondante (donc le tableau 2 ou le tableau 3). Pour un homme, je donne le placebo, car la statistique du tableau 2 pour les hommes me conseille de donner le placebo. Pour une femme, je donne aussi le placebo, car la statistique du tableau 3 pour les femmes me dit que c'est préférable. L'information sur le sexe du patient me conduit à changer de prescription. Ce n'est pas absurde : l'apport de nouvelles informations justifie souvent de changer ses choix.

Mais à y regarder de près, ce point de vue met mal à l'aise et ne semble pas rationnel. C'est ce qu'exprime le second point de vue.

Point de vue 2. La conclusion du point de vue 1 est absurde. Un patient est soit un homme, soit une femme, et que ce soit l'un ou l'autre quand je connais le sexe du patient la consigne déduite est la même : donner le placebo. Ma décision ne dépend pas du résultat de la question : « S'agit-il d'un homme ou d'une femme ? » Il en résulte donc, de toute évidence, que même quand j'ignore le sexe du patient, je dois donner le placebo et ne pas

4

Yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	20	20	50 %
Placebo	16	24	40 %

5

Yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	20	20	50 %
Placebo	16	24	40 %

6

H.yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	19	13	59,375 %
Placebo	6	2	75 %

7

H.yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	17	11	60,71 %
Placebo	8	4	66,66 %

8

F.yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	1	7	12,5 %
Placebo	10	22	31,25 %

9

F.yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	3	9	25 %
Placebo	8	20	28,57 %

tenir compte de la statistique générale qui ne sert plus à rien dès l'instant où je dispose des deux statistiques particulières. On se trouve dans un cas assez usuel : la connaissance de nouvelles données (ici les statistiques particulières concernant séparément les hommes et les femmes) change la conclusion que je faisais avant d'en disposer. Ce changement est général et n'est pas dû à ma connaissance du sexe du patient, mais à ma connaissance des tableaux 2 et 3. Il n'y a rien d'irrationnel à changer d'avis quand on est mieux informé, certes, mais ici le changement doit conduire à oublier le tableau 1 qui ne sert plus à rien quand on a les deux autres.

Notons que, pour tirer une règle de données statistiques, on exige en général d'avoir des effectifs totaux plus élevés que ceux qui apparaissent dans nos tableaux. Mais cela est sans importance pour tout ce que nous venons de dire et pour tout ce que nous dirons plus loin, car on peut très bien imaginer que tous les nombres de nos tableaux sont multipliés par 100 ou même 10 000 et que les déductions que nous faisons des statistiques des tableaux sont de ce fait parfaitement sûres.

Le point de vue 2 est confirmé par un argument formel : de (A ⇒ B) et (non-A ⇒ B), on déduit B. C'est ce qu'on nomme parfois le « principe de la chose sûre » : si, lorsque A est vrai, j'en déduis B, et que, lorsque A est faux, j'en déduis aussi B, alors c'est que B est toujours vrai.

Le point de vue 2 correspond à l'analyse classique du paradoxe et semble nous en libérer. On l'exprime parfois en disant qu'il faut prendre des précautions pour agréger des données. Je croyais satisfaisant ce point de vue, quand j'ai pris conscience d'une critique qui en montre l'absurdité.

Plus rien ne va !

Critique du point de vue 2. Imaginons que je dispose aussi d'une statistique sur les patients aux yeux clairs et d'une statistique sur les patients aux yeux foncés, et qu'elles mènent chacune à la conclusion que le médicament est plus efficace que le placebo.

Une telle situation est tout à fait possible et les données que nous proposons pour les tableaux 4 et 5 sont compatibles avec les trois premiers tableaux. C'est ce que démontrent les tableaux suivants 6, 7, 8 et 9 ; pour chacune des quatre catégories « Hommes aux yeux clairs », « Hommes aux yeux foncés », « Femmes aux yeux clairs », « Femmes aux yeux foncés », ils indiquent les taux de réussite respectifs du médicament et du placebo.

On vérifiera sans peine que les tableaux 6 et 7 redonnent 2, que 8 et 9 redonnent 3, que 6 et 8 redonnent 4, et enfin que 7 et 9 redonnent 5. Une situation comme celle envisagée ici est parfaitement possible.

Avec ces tableaux, le point de vue 2 est insoutenable. En effet, si j'adopte ce point de vue et que je l'applique non pas à la distinction homme-femme, mais à la distinction yeux clairs-yeux foncés, je dois conclure qu'il faut prescrire le médicament. Le point de vue 2 conduit donc, selon qu'on s'attache au sexe des patients ou à la couleur de leurs yeux, à des prescriptions contradictoires. Dans le cas décrit par l'ensemble des tableaux présentés, tous parfaitement compatibles, le point de vue 2 conduit à deux conclusions opposées ; il serait irrationnel de le défendre. En résumé, le point de vue 1 est clairement absurde, et le point de vue 2 aussi car il conduit à des prescriptions contradictoires. Comment sortir du paradoxe ?

Il semble inévitable d'opter pour un point de vue prudent et d'admettre que sans données plus précises, aucune prescription ne sera sûre. Ce point de vue 3, expliqué ci-dessous, ne satisfera pas le médecin qui n'a pas la possibilité de différer

3. Le «Double-Simpson»

Ce jeu de données, que nous nommons le « Double-Simpson », a été découvert par Jean-François Colonna le 20 mars 2013 en menant des calculs intensifs. Il a les propriétés paradoxales suivantes.

– Dans le tableau 1, on considère une population de 183 personnes séparées en deux groupes complémentaires dont l'un prend le médicament et l'autre un placebo. Le taux de guérison est meilleur pour le groupe prenant le médicament qui semble donc devoir être prescrit aux malades.

– Quand on sépare les hommes et les femmes (tableaux 2 et 3), pour chaque catégorie le placebo est meilleur que le médicament. C'est le paradoxe de Simpson simple.

– Quand on considère séparément les personnes aux yeux clairs et les personnes aux yeux foncés (tableaux 4 et 5), le médicament est à nouveau meilleur.

– De plus, et ce fut le point le plus délicat à satisfaire, les données des tableaux 1-5 résultent de deux séries de tableaux, soit A6-A7-A8-A9, soit B6-B7-B8-B9, aux propriétés opposées. Pour chacune des quatre catégories (« Hommes aux yeux clairs », « Hommes aux yeux foncés », « Femmes aux yeux clairs », « Femmes aux yeux foncés »), selon qu'on lit les tableaux A ou B dans la même catégorie, on trouve que le médicament surpasse le placebo ou l'inverse, mais on n'a jamais le même résultat pour A et B.

Cette divergence systématique entre les données A et B montre que quelqu'un disposant des tableaux 1-5 ne peut rien conclure pour aucune personne quelle que soit sa catégorie (Homme, Femme, Yeux clairs, Yeux foncés).

1

Général	Guéri	Non guéri	Taux de guérison
Médicament	47	42	52,80 %
Placebo	38	56	40,42 %

2

Hommes	Guéri	Non guéri	Taux de guérison
Médicament	40	26	60,60 %
Placebo	16	9	64 %

3

Femmes	Guéri	Non guéri	Taux de guérison
Médicament	7	16	30,43 %
Placebo	22	47	31,88 %

4

Yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	24	19	55,81 %
Placebo	21	28	42,85 %

5

Yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	23	23	50 %
Placebo	17	28	37,77 %

A6

H. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	21	13	61,76 %
Placebo	8	5	61,53 %

A7

H. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	19	13	59,37 %
Placebo	8	4	66,66 %

A8

F. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	3	6	33,33 %
Placebo	13	23	36,11 %

A9

F. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	4	10	28,57 %
Placebo	9	24	27,27 %

B6

H. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	21	14	60 %
Placebo	9	4	69,23 %

B7

H. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	19	12	61,29 %
Placebo	7	5	58,33 %

B8

F. yeux clairs	Guéri	Non guéri	Taux de guérison
Médicament	3	5	37,50 %
Placebo	12	24	33,33 %

B9

F. yeux foncés	Guéri	Non guéri	Taux de guérison
Médicament	4	11	26,66 %
Placebo	10	23	30,30 %

pls_429_p080_085_delahaye.indd 83 07/06/13 18:21

son choix, mais l'éventualité de cas tels que ceux décrits ci-dessus, et encore plus tels que celui décrit à la figure 4, force cette conclusion attentiste.

Une attitude prudente

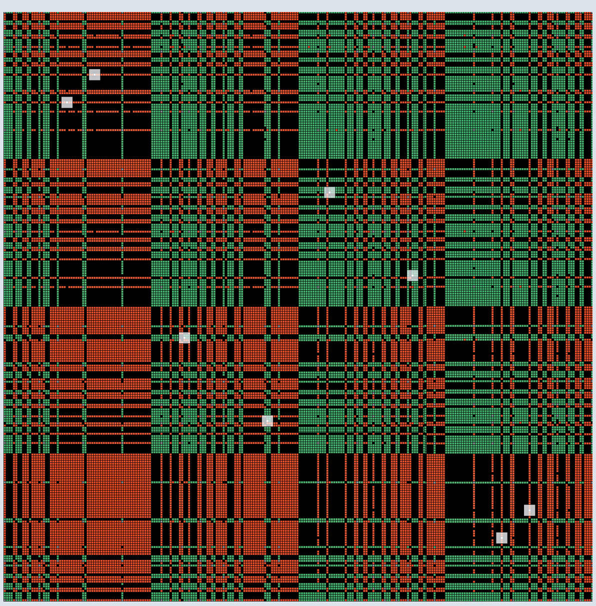
Point de vue 3. Dans une situation où seules les données des tableaux 1, 2 et 3 sont disponibles, aucune prescription ne s'impose absolument, du fait de l'existence possible d'une partition contradictoire comme celle donnée par les tableaux 4 et 5. Celui qui dispose des données des tableaux 6, 7, 8 et 9, qui indiquent pour chaque catégorie précise (sexe connu et couleur des yeux connue) si le médicament est meilleur que le placebo pourra en revanche prescrire un traitement à un patient dont on connaît le sexe et la couleur des yeux : il se référera au tableau le concernant. Ce jugement prudent est confirmé par deux arguments.

Argument 1. Quand on dispose de données analogues à celles du tableau 1 (indiquant que « le taux de guérison est meilleur avec le médicament qu'avec le placebo ») et que les effectifs ne sont pas trop petits (supérieurs à 3), il est toujours possible de faire une partition des patients en deux catégories complémentaires de façon à avoir des tableaux 2 et 3 produisant des conclusions opposées à celle du tableau 1 (« le placebo est meilleur »).

En effet, composons la catégorie 1 de patients guéris par le placebo (on en garde au moins un pour la catégorie 2) et aucun de la catégorie que le placebo ne guérit pas, auxquels on ajoute tous les patients que le médicament guérit et quelques patients que le médicament ne guérit pas. Ce choix assure que dans la catégorie 1, le taux de réussite du placebo est 100 %, et qu'il ne l'est pas pour le médicament. Dans la catégorie 2, il n'y a aucun patient que le médicament guérit (ils ont tous été mis dans la catégorie 1) et donc le taux de guérison avec le médicament est nul, ce qu'il n'est pas pour le placebo puisqu'on a réservé au moins un patient guéri par le placebo pour la catégorie 2. Aussi bien dans la catégorie 1 que 2, le placebo sera donc strictement meilleur que le médicament. Si les coefficients du tableau 1 sont (2, 2, 2, 3), la méthode décrite donne (2, 1, 1, 0) pour le tableau 1 et (0, 1, 1, 3) pour le tableau 2.

Ce type de partages est certes artificiel, mais il montre que le risque de paradoxe de

4. Visualisation de la fréquence du paradoxe



$27 = 0\times64 + 1\times16 + 2\times4 + 3\times1 = 0123$ en base 4

27	Guéri	Non guéri	Taux de guérison
Médicament	1	3	25 %
Placebo	2	4	33 %

Le placebo est meilleur

$216 = 3\times64 + 1\times16 + 2\times4 + 0\times1 = 3120$ en base 4

216	Guéri	Non guéri	Taux de guérison
Médicament	4	3	57 %
Placebo	2	1	66 %

Le placebo est meilleur

MATRICE SOMME

243	Guéri	Non guéri	Taux de guérison
Médicament	5	6	45,45 %
Placebo	4	5	44,44 %

Le médicament est meilleur

Cette figure représente l'ensemble des cas possibles lorsqu'on considère le nombre de guérisons pour des patients hommes prenant le médicament ou le placebo et pour des patientes femmes prenant le médicament ou le placebo (cela correspond aux tableaux 2 et 3 du texte). On a envisagé tous les cas lorsque les huit éléments des deux tableaux parcourent toutes les valeurs entières entre 1 et 4. Cela fait 4^8 = 65 536 cas différents.

En regroupant les tableaux hommes et femmes d'un cas, on obtient un tableau de synthèse (hommes et femmes réunis) qui donne parfois une situation de type « paradoxe de Simpson » : les tableaux hommes et femmes vont dans le même sens (le médicament est meilleur, ou le placebo est meilleur), mais le tableau synthétique donne une conclusion opposée.

Chacun des 65 536 cas est représenté par un pixel. Les quatre coefficients du tableau pour les hommes (première colonne : a, b ; deuxième colonne : c, d) fournissent l'abscisse du pixel en considérant le nombre entier dont l'écriture en base 4 a pour chiffres $(a-1)$, $(b-1)$, $(c-1)$, $(d-1)$. Les quatre coefficients du tableau pour les femmes donnent, selon le même codage, l'ordonnée du pixel. La couleur du pixel est choisie selon les règles suivantes.

– Si on a un cas « normal » (les tableaux hommes et femmes donnent une même conclusion, qui est aussi celle du tableau de synthèse), le pixel est vert ou marron selon que le placebo l'emporte dans le cas général ou non.

– Si on a un cas sans grand intérêt (les tableaux hommes et femmes se contredisent, ou il y a égalité entre les succès du médicament et du placebo), le pixel est noir.

– Si on a un cas de paradoxe de Simpson (le tableau de synthèse s'oppose aux deux tableaux hommes et femmes qui vont dans le même sens), le pixel est clair, entouré d'un halo pour aider à le repérer. Il n'y a que huit cas, correspondant aux huit pixels (27, 216), (39,228), (78,114), (114,78), (141,177), (177, 141), (216, 27) et (228, 39), le premier étant détaillé ci-contre.

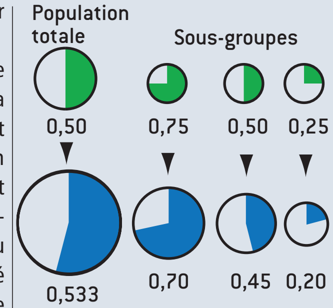
Cette représentation, imaginée et réalisée par Jean-François Colonna, montre qu'heureusement le paradoxe de Simpson se produit assez rarement.

5. Le paradoxe de Simpson et l'émergence de la coopération

Dans la compétition entre organismes biologiques, le maintien des « biens communs » est une question importante. Lorsqu'il y a un producteur de ce bien commun et un consommateur (un organisme qui en bénéficie sans le produire), les consommateurs ont tendance à proliférer aux dépens des producteurs. L'oxygène de l'air est un exemple de bien commun produit seulement par certains organismes et consommé par d'autres qui ne le produisent pas. En 2009, des chercheurs de l'Université Rockefeller à New York (John Chuang, Olivier Rivoire, Stanislas Leibler) ont créé et étudié une telle situation avec deux souches de la bactérie *Escherichia coli*. L'une des souches produisait un antibiotique utile aux deux souches et l'autre en profitait sans le produire. Les chercheurs ont composé plusieurs groupes de bactéries, chaque groupe comportant les deux types de bactéries. Une situation apparemment paradoxale est apparue : les non-producteurs croissaient plus vite dans chaque groupe (ce qui n'est pas une surprise) bien qu'au total, les non-producteurs voyaient leur effectif global décroître en proportion. Il s'agissait d'un phénomène de type paradoxe de Simpson.

Pour un biologiste, ce système est un exemple frappant de conflit entre niveaux de sélection. Les bactéries productrices du bien commun sont les bénéficiaires de l'ensemble du système quand on les considère comme un tout, alors qu'à un niveau individuel (celui auquel opère la sélection), elles sont désavantagées, puisque dans chaque groupe leur proportion diminue.

Grâce aux effets du paradoxe de Simpson, un trait qui bénéficie à la population considérée comme un tout peut ainsi se trouver sélectionné, bien qu'à un niveau individuel le trait soit désavantageux. La chose est étonnante : ce qui est mauvais au niveau individuel se trouve au total favorisé par l'effet mécanique d'un paradoxe de Simpson. La réalisation concrète de l'expérience par les chercheurs démontre que cet effet sélectif paradoxal n'est pas seulement théorique, mais doit être pris en compte par les spécialistes de l'évolution. Cette dynamique doit être envisagée comme mécanisme de sélection de traits individuels favorables à la coopération et à l'altruisme. Utiliser le paradoxe de Simpson est l'une des ruses que la sélection naturelle mettrait en œuvre pour favoriser les traits coopératifs et faire ainsi émerger et prospérer des entités collectives d'individus coopérateurs et solidaires.



En vert, la proportion initiale de bactéries de type 1.
En bleu, la proportion finale de bactéries de type 1.

Simpson est toujours présent : tout tableau composé de données pas trop petites est susceptible d'être séparé en deux tableaux créant un paradoxe de Simpson.

Argument 2. Il existe des situations que l'on nommera « Double-Simpson », dont la première a été découverte par Jean-François Colonna, du Centre de mathématiques appliquées de l'École polytechnique, où :

– Le tableau 1 suggère que le médicament est meilleur que le placebo.

– Les tableaux 2 et 3 suggèrent que le placebo est meilleur que le médicament.

– Les tableaux 4 et 5 suggèrent que le médicament est meilleur que le placebo.

– Il existe deux jeux de données (tableaux 6A-9A et tableaux 6B-9B) compatibles avec les tableaux 1-5 et tels que pour chaque catégorie précise (il y en a quatre), ce que laissent supposer les données A est l'inverse de ce que suggèrent les données B *(voir l'encadré 3)*.

Cette possibilité de jeux A et B conduisant à des prescriptions opposées pour chacune des quatre catégories n'était pas connue avant les expériences numériques de J.-F. Colonna. Elle établit définitivement que dans certaines situations, celui qui veut des certitudes absolues ne peut agir à cause du paradoxe de Simpson.

Heureusement, dans les cas que nous avons envisagés, celui qui est prêt à prendre le risque de se tromper (en médecine le prescripteur est souvent dans une telle situation) réussira à prendre des décisions à peu près sûres. On montre en effet que celui qui connaît un tableau comme notre tableau 1 et le sépare en deux catégories complémentaires ne tombera sur des résultats opposés (tels les tableaux 2 et 3) qu'avec une probabilité de 1,92 %. De même, en fusionnant deux tableaux donnant des indications dans le même sens, la probabilité que le tableau résultant contredise les deux premiers est faible. De même encore, la probabilité que trois tableaux, dont l'un est la somme des deux autres, présentent un paradoxe de Simpson est faible.

La conclusion est que même si les cas où le paradoxe de Simpson se produit sont vraiment gênants (c'est ce qu'établit le « Double-Simpson »), ils sont heureusement assez rares. On peut donc prendre le risque... de ne pas y penser ! ■

■ L'AUTEUR

J.-P. DELAHAYE est professeur à l'Université de Lille et chercheur au Laboratoire d'informatique fondamentale de Lille (LIFL).

■ BIBLIOGRAPHIE

J. Pearl, **Simpson's paradox : An anatomy**, 2011 : http://escholarship.org/uc/item/3s62r0d6

H. Goltz et M. Smith, **Yule-Simpson paradox in research**, *Practical Assessment, Research & Evaluation*, vol. 15, 2010 : http://pareonline.net/pdf/v15n15.pdf

J. Chuang *et al.*, **Simpson's paradox in a synthetic microbial system**, *Science*, vol. 323, pp. 272-275, 2009.

M. Pavlides et M. Perlmen, **How likely is Simpson's paradox ?** *The American Statistician*, 2009 : http://bit.ly/10xToug