

L'étonnement vient de ce que l'on croit que $[a/b < c/d \text{ et } a'/b' < c'/d']$ entraîne $A/B < C/D$, ou encore avec seulement les petites lettres: $[a/b < c/d \text{ et } a'/b' < c'/d']$ 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 propor-

tion 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 décliné, 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

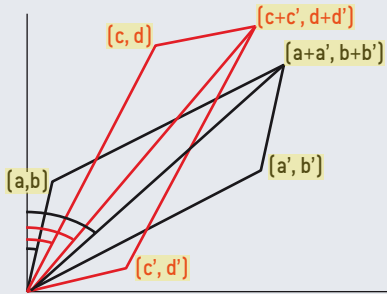


Rue en Suède



Limitation de vitesse

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

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 $(A \Rightarrow B)$ et $(\text{non-}A \Rightarrow B)$, 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

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

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

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

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

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

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