Compléments sur le Khideux

in Documentation méthodologique de Trideux

Philippe Cibois

Mise à jour juillet 2009

Risque de première espèce, risque de deuxième espèce.

C'est un vocabulaire que les statisticiens emploient, même quand ils parlent à des sociologues : ce n'est pas indispensable de parler de risque de première espèce et de deuxième espèce. Il s'agit de la généralisation du cas particulier plus simple : risque que le Khideux soit grand du simple fait du hasard d'échantillonnage, contre risque d'être trop prudent et de ne pas voir une liaison. C'est la réalité de ces risques qu'il faut faire comprendre, non faire apprendre une terminologie abstraite.

Quand je jette 5 pièces de monnaie sur la table et que je dis qu'elles tombent toutes sur le même côté, les étudiants hésitent entre deux interprétations : croire au "miracle" du hasard ou croire que je truque la réalité. Ils sont pris entre le risque d'admettre comme vrai un évènement peu probable (proche d'ailleurs de 5%) et celui de ne pas voir que le prestidigitateur a un "truc". Comme ils ne croient pas au hasard, comme tout un chacun car la propension à y croire est socialement peu répandue, ils sont rassurés quand l'étudiant témoin de l'expérience révèle que les 5 pièces ne sont pas tombées du même côté.

Pour que les étudiants comprennent ce que sont les risques sous-jacents à une décision, il faut enraciner le risque statistique dans celui des décisions de la vie courante. Par exemple dans l'action de tous les jours nous faisons des choix raisonnables dès que nous nous sortons du choix au hasard (50 contre 50). Avoir 2 chances sur 3 d'avoir raison est souvent considéré dans l'action pratique comme raisonnable : c'étaient d'ailleurs les chances que Napoléon exigeait pour engager une bataille selon Joseph Caillaux (règle qu'il appliqua pour son propre compte en refusant d'engager les hostilités avec l'Allemagne à Agadir en 1911 selon ses Mémoires). On comprend que si on était plus exigeant pour engager une bataille, on ne livrerait que des combats gagnés d'avance (du style de l'embuscade), ce qui n'est pas toujours possible. Pour que les généraux acceptent de s'affronter (comme dans les grandes batailles des guerres napoléoniennes), il faut qu'ils estiment avoir des chances raisonnables de gagner et ce ne peuvent être des probabilités trop fortes.

Dans les termes de la statistique, travailler au seuil de Napoléon c'est prendre un seuil de 33%. Dans le cas d'un intervalle de confiance autour d'une moyenne, c'est prendre un seul écart-type à droite ou à gauche. Dans une étude exploratoire, ce n'est pas une mauvaise stratégie.

Calcul du Khideux

En donnant une formule sans explication (observé moins théorique au carré divisé par théorique) on incite à la recette de cuisine. D'autant qu'au lieu de reprendre une disposition en tableau de l'effectif d'indépendance puis des écarts à l'indépendance, puis des contributions au Khideux on fait souvent calculer à la suite les différentes contributions comme si la structure tabulaire n'existait plus.

Le rapport "écart / théorique" présenté comme étant un écart relatif, peut aider à faire comprendre le Khideux. Cet indice a été présenté une première fois par J .Bertin (*Sémiologie graphique*, Paris, Mouton, 1967 : 225) sous le non de rapport ou de pourcentage de distorsion, puis repris sous le nom de "khi" (J. Bertin, La graphique et le traitement graphique de l'information, Paris, Flammarion, 1977 : 65). H. Rouanet .(H.Rouanet et al., *Statistique en sciences humaines : procédures naturelles, Paris*, Dunod, 1987 : 158) reprend cette indice sous le nom de "taux de liaison".

On peut faciliter la compréhension du Khideux en utilisant l'indicateur "écart relatif" : il s'agit d'un coefficient de pondération que l'on multiplie par l'écart à l'indépendance et qui nous donne la contribution au Khideux :

- si l'écart à l'indépendance est faible par rapport au théorique, cela signifie que la case apporte peu d'information et la contribution résultante au Khideux sera faible parce qu'inférieure à l'écart à l'indépendance,

- si l'écart à l'indépendance est égal au théorique, l'indice sera égal à l'unité et la contribution au Khideux sera identique à l'écart à l'indépendance,

- enfin, si l'indice est supérieur à un, cela signifie que l'écart à l'indépendance est plus grand que l'effectif théorique et le coefficient de pondération fera que la contribution au Khideux sera plus forte que l'écart à l'indépendance, ce qui manifeste que cette case apporte beaucoup d'information.

On considère ainsi la contribution au Khideux comme un indicateur de l'information apportée par une case et on relie le Khideux d'une manière simple avec la théorie de l'information qui lui est liée. De plus cela permet de montrer que le Khideux est homogène à un effectif d'écart, non à un carré d'écart, que c'est un écart pondéré.

On peut facilement le vérifier en voyant que dans un tableau quelconque, le Khideux total est toujours du même ordre de grandeur que la somme des écarts positifs à l'indépendance. Les Khideux sont des écarts pondérés, non des nombres magiques. Quand le Khideux est fort, cela signifie que les écarts sont forts, et donc que la situation d'indépendance de la population de référence devient très improbable.

Limites et précautions

On tombe souvent dans les limites et précautions associées à l'usage du Khideux, non plus dans la recette de cuisine, mais dans le rite magique avec des "interdictions".

Pour comprendre ce qui se passe regardons d'abord comme les avis sont divers, ce qui est quand même rare pour un sujet qui touche aux mathématiques :

Recettes de Khideux

When the expected frequencies in a fourfold table are small the approximation obtained by the usual X^2 tables will be improved if we calculate X^2 from the first expression [Cramér désigne là le khi-deux usuel], and reduce the absolute value of each difference [observés - théoriques] by ½ before squaring. This is known as Yates' correction.

Harald Cramér, *Mathematical methods of statistics*, PUP, 1946, p.445

L'application du critère de distance n'est valable que si les effectifs théoriques des différentes modalités ne sont pas trop petits. En pratique, on considère généralement qu'ils doivent être au moins égaux à 4 ou 5.

G.Calot, *Cours de statistique descriptive*, Dunod, 1965, p.180

If some of the Eij (expected number of observations) are small, the approximation may be very poor. Cochran (1952) states that if any Eij is less than 1 or il more than 20% of the Eij are less than 5, the approximation may be poor. This seems to be overly conservative (...). If r and c are not too small, I feel that the Eij may be as small as 1.0 without endangering the validity of the test.

W.J. Conover, *Practical Nonparametric Statistics*, J.Wiley, 1980, p.156

Conditions de validité : les effectifs théoriques doivent être tous supérieurs ou égaux à 10 (ou à la rigueur 5)

Philippe Lazar et Daniel Schwartz, *Eléments de probabilités et statistique*, Flammarion, 1987, p.89

Mêmes auteurs, édition antérieure (1964) : la méthode n'est valable que si les effectifs calculés égalent ou dépassent 5 (10 pour plus de rigueur) p.70

Effectif total supérieur ou égal à 30, tous les théoriques supérieurs ou égal à 1, et 80% d'entre eux supérieurs ou égaux à 5

Jean-Jacques Droesbeke, *Eléments de statistiques*, Ellipses,1988, p.354

En termes de la statistique X^2, la règle pratique de validité de l'approximation s'exprime ainsi : les effectifs théoriques doivent tous être supérieurs à 5. L'approximation est grandement améliorée si l'on effectue la correction de continuité.

H.Rouanet, J.M.Bernard, B.Leroux, *Analyse inductive des données*, Dunod, 1990, p.154

Cochran (1952,1954) recommends that a minimum expected cell frequencey as low as 1 be allowed if no more than 20% of cells have expected frequencies less than 5. If X^2 has less than 30 degrees of freedom and the minimum expected frequency is 2 or more, Cochran (1954) states that the use of the ordinary X^2 tables is usually adequate.

Wayne W.Daniel, *Applied nonparametric statistics*, PWS-KENT, 1990, p.185

On devrait avoir les théoriques supérieur ou égal à 1 dans chaque classe (certains auteurs plus sévères dites plus grand ou égal à 5)

Th. H. Wonnacott et R. J. Wonnacot, *Statistique*, Economica, 1991, traduction d'un manuel célèbre dans les pays anglo-saxons, p.622

Most (e.g. more than 80 percent) of the estimated expected frequencies are larger than 5.S. Lewis-Beck (ed.) *Basic Statistics*, Sage, 1993, p.171

On voit au fil du temps un passage (qui ne s'impose pas pour tous) de l'influence de Fisher à celle de Cochran : cependant la variabilité des citations devrait ôter l'idée qu'il y a une règle qui s'impose alors qu'il doit y avoir simplement une attitude de prudence.

Dans Fisher on trouve seulement mention de la règle "it is desirable that the number expected should in no group be less than 5" (dans son chapitre "Tests of goodness of fit, independance and homogeneity ; with table of X^2" extrait de la réédition de 1990 de *Statistical Methods, Experimental Design and Scientific Inference*, Oxford OUP.p.84)

Les recettes ont changé : aujourd'hui la règle de Cochran est parfois imposée d'une façon aussi impérative que la règle de Fisher auparavant, (de la même manière que les pédiatres ont impérativement ordonné dans les années 70 que les bébés devaient dormir sur le ventre alors qu'aujourd'hui, tests statistiques relatifs à la MSN à l'appui, ils disent qu'ils doivent dormir sur le dos). Les règles de Fisher et de Cochran datent d'avant l'époque de l'informatique : il est possible aujourd'hui d'inspecter à vue les contributions au Khideux de chaque case.

Correction de Yates :

Fisher en parle : voici son texte (Fisher 1990, p. 92-94)

21.01 Yates' Correction for Continuity

The distribution of X^2, tabulated as in Table III, is a continuous distribution. The distribution of frequencies must, however, always be discontinuous. Consequently, the use of X^2 in the comparison of observed with expected frequencies can only be of approximate accuracy, the continuous distribution being in fact the limit towards which the true discontinuous distribution tends as the sample is made ever larger. It was in order to avoid the irregularities produced by small numbers that we have stipulated above that in no group shall the expected number be less than five. This safeguard generally ensures that the number of possible sets of observations shall be large, each occurring with only a small frequency, so giving to X^2 a distribution closely simulating the continuous distribution of the table.

1. A case of special interest arises, however, when there is only I degree of freedom, and when the value of X^2 can, consequently, be calculated from the number observed in a single class. If the number in this class is small, e.g. 3, the probability of this number may be by no means negligible compared with the sum of the probabilities of the more extreme deviations represented by 2, I, or 0 occurrences in the class. If we want to know whether the observed number, 3, is so small as to indicate a significant departure from expectation, we require to know whether the sum of the probabilities of 3, 2, I or 0 together is less than a standard value, such as .05; or, in other words, whether the total probability of obtaining our observed deviation, or any deviation more extreme, is so small that we should be unwilling to ascribe the deviation observed to mere chance.

Our actual problem, therefore, when stated exactly, concerns a limited number of finite probabilities, which in simple cases it may be convenient to calculate directly. The

> Table of X^2, on the other hand, gives the area of the tail of a continuous curve. Inasmuch, however, as this curve supplies a close approximation to the actual distribution, the area between the values of X^2 corresponding to observed frequencies of 3½ and 2½ will be a good approximation to the actual probability of observing 3; and the area of the tail beyond the value of X^2 corresponding to 3½ will be a good approximation to the sum of the probabilities of observing 3 or less. Thus our actual problem will best be resolved by entering the table of X^2, not with the value calculated from the actual frequencies, but with the value it would have if our observed frequencies had been less extreme than they really are each by half a unit. This useful adjustment is due to F. Yates.

La correction de Yates est une correction de continuité : une courbe de probabilité cumulée est, sur les faibles fréquences, en escalier et s'écarte de la courbe théorique qui est régulière. Pour pallier cette difficulté, on ajoute 0,5 pour être plus près de la courbe théorique. Ceci n'a de sens que pour les faibles effectifs (théorique inférieur à 5 dit Fisher) : pour les théoriques plus grand que 5, cela ne change pratiquement rien.

Certains auteurs affirment que la correction de Yates doit être utilisée quand les effectifs théoriques sont inférieurs à 500. Je croyais cette erreur imputable à une rumeur du milieu statisticien, il n'en est rien : dans la troisième édition (1949) des Statistical Tables for Biological, Agricultural and Medical Research de Ronald Fisher et Frank Yates (Edinburgh, Oliver and Boyd), il est dit p.4 que la correction de continuité "should be applied when the smallest expectation is less than 500". On notera cependant que dans l'exemple numérique qui suit, le plus petit effectif théorique est égal à 7,4. Y aurait-il divergence entre Fisher et Yates ? Il faudrait remonter au texte initial de Yates.

Conclusion

Ces quelques réflexions suggèrent que l'enseignement du Khideux doit :

1) tenir compte du fait que l'enseignement est fait à de futurs sociologues dont c'est souvent le premier contact avec l'univers des tests. Plutôt que d'utiliser le vocabulaire des statisticiens (hypothèse nulle, risques de première ou de deuxième espèce), il faut en faire comprendre la substance.

2) la contribution au Khideux est homogène à un écart, c'est un écart pondéré : cette manière de regarder le Khideux fait qu'il n'est plus un nombre magique, mais qu'il s'agit de sommer des effectifs pondérés.

3) les précautions à prendre pour l'usage du Khideux sont modifiées par l'utilisation de l'informatique : l'inspection à vue des contributions permet d'éviter à l'utilisateur de prendre les règles de précautions comme des recettes magiques.