



Un modèle d'accord entre observateurs sur une échelle nominale

Jean-Marie Tricot

► To cite this version:

Jean-Marie Tricot. Un modèle d'accord entre observateurs sur une échelle nominale. Comptes Rendus Mathématiques, Rep. Acad. Sci. Canada, 1991, 13 (4), pp.146-150. hal-00989160

HAL Id: hal-00989160

<https://hal.science/hal-00989160>

Submitted on 15 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Volume XIII, No. 4, August 1991 août

**COMPTES RENDUS
MATHÉMATIQUES**

DE L'ACADEMIE DES SCIENCES

MATHEMATICAL REPORTS

OF THE ACADEMY OF SCIENCE



**The Royal Society of Canada
La Société royale du Canada**

UN MODELE D'ACCORD ENTRE OBSERVATEURS
SUR UNE ECHELLE NOMINALE

Jean-Marie Tricot

Presented by D.A.S. Fraser, F.R.S.C.

Abstract

By an appropriate hypothesis testing in a simulated example, one can show that Cohen's kappa can sometimes take values whose interpretation may be difficult. A new statistics measuring interobserver agreement is derived which does not involve pairwise considerations. This statistics is compared with kappa's measure and can be interpreted as a sample variance. Some properties relevant to Partition's Theory show that its values are ordered according to adequate agreement criteria.

1. Introduction et notations

Dans une approche nouvelle, on désire étudier la concordance entre d observateurs évaluant n sujets sur une échelle qualitative nominale, d'un point de vue global sans faire intervenir la notion de couple d'observateurs ou de corrélation. On indiquera quels liens existent entre le coefficient kappa introduit par [Co] et le nouveau coefficient proposé, en se plaçant dans le cas non pondéré. On fera appel à la notion de partition reprenant ainsi l'approche de [Tr].

On suppose que les évaluations (ou catégories) sont les éléments d'un ensemble fini d'évaluations E dans N sans structure de cardinal k . Soit $(i_1, \dots, i_{m_i})$, $i=1, n$, le sous-ensemble des m_i évaluations de E données au sujet i . On note z_q^i , $q=1, m_i$, le nombre d'observateurs ayant porté l'évaluation i_q sur le sujet i . La suite $p^i = (z_1^i, \dots, z_{m_i}^i)$ est un élément de l'ensemble D des partitions de l'entier d . On posera $D^* = D \setminus \{d\}$.

2. Les limites de l'interprétation des valeurs de la mesure kappa

[LK] réalisent des tests de comparaison entre différentes statistiques kappa définies sur différents tableaux de données et pour différents systèmes de pondérations; ou des tests de signification sur ces mêmes statistiques. Les interprétations en termes de degrés comparés d'accord interobservateur sont évidemment liées à l'interprétation des différentes valeurs possibles du coefficient $\kappa = (\pi_0 - \pi_c) / (1 - \pi_c)$. Voilà pourquoi les auteurs, cités plus haut, nous donnent quelques "points de repères" sur la correspondance entre certaines plages d'estimations de κ et "l'intensité de l'accord".

A titre d'illustration, reprenons le test de l'hypothèse paramétrique $\kappa = 0$ en appliquant les formules de [Sc] au cas où $\text{Card}(E) = k \geq 4$, k pair, $d \geq 4$, d pair, $2n$ divisible par k ; chaque sujet a reçu deux évaluations distinctes, chacune de la part de la moitié des observateurs; enfin, $\forall q_1, q_2, q_1 \neq q_2, \sum_i z_{q_1}^i = \sum_i z_{q_2}^i$. D'où:

$$\hat{\kappa} = \frac{k(d-2) - 2(d-1)}{2(d-1)(k-1)} \geq \frac{d-3}{3(d-1)}$$

$$s.e..(\hat{\kappa}) = \frac{\sqrt{2}}{\sqrt{nd(d-1)(k-1)}} \leq \frac{1}{(d-1)\sqrt{n}}$$

$$\hat{\kappa} / s.e..(\hat{\kappa}) \geq \frac{d-3}{3} \sqrt{n}$$

Par conséquent, la statistique du test de $\kappa = 0$ peut prendre des valeurs arbitrairement grandes avec d ou n . Le rejet de $\kappa = 0$, pour un seuil de signification arbitrairement grand, peut donc se produire dans une situation caractéristique de désaccord.

Ajoutons enfin que si, par exemple, $k=d=15$, on obtient $\hat{\kappa}=.426$, valeur reflétant une intensité de l'accord "moyenne" selon [LK], alors que l'on pourrait aussi parler d'absence d'accord.

3. La mesure kappa exprimée comme un écart à une loi marginale

On a les estimations usuelles de π_0 et π_c :

$$\hat{\pi}_0 = \frac{1}{n} \sum_i \frac{1}{d(d-1)} \sum_q z_q^i (z_q^i - 1)$$

$$\hat{\pi}_c = \sum_q \left(\frac{1}{nd} \sum_i z_q^i \right)^2$$

D'où:

$$\hat{\kappa} = \frac{\hat{\pi}_0 - \hat{\pi}_c}{1 - \hat{\pi}_c} = \frac{d}{n(d-1)(1 - \hat{\pi}_c)} \sum_i \sum_q \left[\frac{z_q^i}{d} - \frac{\bar{z}_q}{d} \right]^2 - \frac{1}{d-1}$$

où $\bar{z}_q = (1/n) \sum_i z_q^i$.

$\hat{\kappa}$ est donc une fonction croissante des écarts des différentes proportions multinomiales z_q^i/d d'une ligne du tableau (z_q^i) , aux proportions marginales $\bar{z}_q/d$ lorsque celles-ci sont fixées. La forme d'une telle expression prise par $\hat{\kappa}$ est tout à fait cohérente avec l'hypothèse de l'absence d'accord formulée par [F1]: les évaluations sont attribuées suivant des probabilités invariantes d'un sujet à l'autre, estimées par $\bar{z}_q/d$.

4. Notion de désaccord et critères du degré d'accord

L'illustration énoncée plus haut, souligne l'avantage de définir le désaccord sur un sujet donné, comme la présence d'une répartition des observateurs en groupes d'égale importance numérique: la somme des écarts des z_q^i à d/m_i ($p^i \in D^*$), évaluée par $\sum_q (z_q^i - d/m_i)^2$ va donc être un 1^{er} critère du degré de l'accord. Le 2^{ème} critère qui s'impose est le nombre m_i . Comme synthèse des deux critères précédents, le critère de la variance des composantes de p^i ,

$$\text{Var}(p^i) = (1/m_i) \sum_{q=1}^m (z_q^i - d/m_i)^2$$

répond à de bonnes propriétés pour mesurer le degré de l'accord. En effet, on a les résultats généraux suivants sur l'ensemble D :

Proposition 1. Soit $p=(z_1, \dots, z_m) \in D$. $\text{Var}(p)$ atteint sa valeur minimum si et seulement si $\forall q \in \{1, \dots, m\}, z_q = E p$.

Proposition 2. Soient $p=(z_1, \dots, z_{q'}, \dots, z_{q''}, \dots, z_m) \in D^*$ et $p'=(z_1, \dots, z_{q'-1}, \dots, z_{q''+1}, \dots, z_m)$ telle que $1 \leq q' < q'' \leq m$ et $2 \leq z_{q'} \leq z_{q''} \leq d-1$. Alors: $\text{Var}(p) < \text{Var}(p')$.

Proposition 3. Soient $p=(1^{m-1}, d-m+1) \in D^*$ et $p'=(1^{m-2}, d-m+2)$ telle que $m \geq 3$ et $d \geq 3$. Alors: $\text{Var}(p) < \text{Var}(p')$.

Corollaire. $\text{Var}(p)$ atteint sa valeur maximum si et seulement si $p=(1, d-1)$.

En conclusion, on admettra donc que, si p^i et $p^{i'}$, éléments de D^* , sont des partitions observées sur les sujets i et i' , l'intensité de l'accord a été plus grande dans le cas de i' que dans le cas de i si $\text{Var}(p^i) < \text{Var}(p^{i'})$.

5. Une nouvelle mesure de concordance

Remarque 1. Par convention, on pourra caractériser l'accord parfait à l'aide de la quantité $d^2/4$ qui représente la variance entre les éléments du couple $(0, d)$.

Remarque 2. L'échelle des valeurs de $\text{Var}(p)$, lorsque p décrit D^* , est conservée si l'on ajoute à chacune de ces variances une quantité inférieure à l'écart minimum entre ces différentes variances: $\forall m \geq 2, d^2/(4m^{12d})$ est une quantité adéquate.

En résumé, la statistique $(1/m_i) \sum_{q=1}^{m_i} (z_q^i - d/m_i)^2 + d^2/(4m_i^{12d})$ mesure bien l'accord interobservateur sur le sujet i lorsque $m_i \geq 2$ puisqu'elle préserve l'ordre des valeurs de $(1/m_i) \sum_{q=1}^{m_i} (z_q^i - d/m_i)^2$. Elle atteint

son maximum lorsque $m_i=1$ (accord parfait). Elle approxime la variance de $(z_1^i, \dots, z_m^i)$ lorsque $m_i \geq 2$, et est égale à la variance de $(0, d)$ lorsque $m_i=1$. Elle est comprise entre $d^2/(4d^{12d})$ et $d^2/4$. On définit donc, par normalisation, comme nouvelle mesure de concordance, la moyenne statistique sur les sujets:

$$V = \frac{4}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{q=1}^{m_i} \left[\frac{z_q^i}{d} - \frac{1}{m_i} \right]^2 + \frac{1}{n} \sum_{i=1}^n m_i^{-12d}$$

V est une fonction croissante des écarts des proportions multinomiales z_q^i/d ($z_q^i \neq 0$) d'une ligne du tableau des z_q^i à une distribution uniforme sur les évaluations prises en compte.

Références

- [Co] J. Cohen, A coefficient of agreement for nominal scales, *Educ. and Psychol. Measurement*, 20 (1960), 37-46.
- [Fl] J.L. Fleiss, Measuring nominal scale agreement among many raters, *Psychological Bulletin*, 76 (1971), 378-382.
- [LK] J.R. Landis and G.G. Koch, The measurement of observer agreement for categorical data, *Biometrics*, 33 (1977), 159-174.
- [Sc] H.J.A. Schouten, Measuring pairwise agreement among many observers. II. Some improvements and additions, *Biom. J.*, 24 (1982), no 5, 497-504.
- [Tr] J.M. Tricot, *Méthode des Réseaux en Analyse de Données, Application à l'Analyse de Concordance*, Thèse de doctorat, Imprimerie Nationale, Genève, 1990.

CONTENTS

F. STREIT	
On the characteristic functions of the Kotz type distributions	121
K.G. NOLTE	
Universal appearance of periodic orbits in the Lorentz equations	125
G. DIAZ et M. MIGNOTTE	
Passage d'une mesure d'approximation à une mesure de transcendance	131
K. LIU	
Quantum central extensions	135
A. GRZYTCZUK and K. INKERI	
On Fermat's equation in the first case	141
J.-M. TRICOT	
Un modèle d'accord entre observateurs sur un échelle nominale	146
M.S. SAMAALI	
Quelques propriétés de la clôture séparable d'un anneau connexe	151
M. MIGNOTTE	
A note on the equation $x^2 - 1 = y^2(z^2 - 1)$	157
B. SZYSZKOWICZ	
Empirical type processes and contiguity	161
D.R. MASSON	
A generalization of Ramanujan's best theorem on continued fractions	167
M. CSORGO and Q.-M. SHAO	
Criteria for limit inferior of small increments of Banach space valued stochastic processes	173
Mailing addresses	179