

Vers une mesure d'accord interjuge prenant en compte le pourcentage d'accord, le hasard et le désaccord

Pier-Olivier Caron

TÉLUQ – Université du Québec
Conférence lors du 40^e congrès annuel de la
Société québécoise de recherche en psychologie : Québec.

24 mars 2018

L'importance de l'accord interjuge

- Observation de comportement
- Validité de la mesure
- Certains comportements sont extrêmement difficiles à classifier

Si les observateurs ne s'entendent pas sur l'occurrence d'une réponse alors aucune science du comportement ne peut se construire

La situation

- Imaginons deux juges qui doivent classifier n objets pouvant être dans k catégories

Juge 1	Juge 2		
	Catégories	1	2
1	50	3	4
2	1	4	13
3	5	15	5

L'accord, le hasard et le désaccord

- Accord
 - Objet classifié identiquement
- Hasard
 - Ce qui est attendu par hasard
- Désaccord
 - Erreur de classification systématique

Juge 1	Juge 2		
	Catégories	1	2
1	50	3	4
2	1	4	13
3	5	15	5

Le pourcentage d'accord

- La somme de la diagonale divisée par le total
- Ne considère que la diagonale
- Ne tient pas compte des fréquences
 - Une **classe** prédominante facile à classer
- Dans cet exemple : 59 %

Juge 1	Juge 2		
	Catégories	1	2
1	50	3	4
2	1	4	13
3	5	15	5

Kappa de Cohen

- Mesure d'accord tenant compte du hasard
 - Ne tient compte que de la diagonale
 - Test : Existe-t-il au moins un accord?
 - Peut être sig., mais impliquer un faible accord

Nouveau test

- Test de χ^2 pondéré pour les fréquences attendues par

où a représente l'accord
$$A = \begin{pmatrix} \frac{a}{k} & \frac{1-a}{k(k-1)} & \dots & \frac{1-a}{k(k-1)} \\ \frac{1-a}{k(k-1)} & \frac{a}{k} & \dots & \frac{1-a}{k(k-1)} \\ \dots & \dots & \dots & \dots \\ \frac{1-a}{k(k-1)} & \frac{1-a}{k(k-1)} & \dots & \frac{a}{k} \end{pmatrix}$$

- On calcule ensuite le χ^2

Méthode

- Algorithme
 1. Générer des accords selon le pourcentage p_a
 2. Générer des désaccords systématiques selon le pourcentage p_d plus exactement $(1-p_a)p_d$
 3. Générer des désaccords aléatoires selon $(1-p_a)(1-p_d)$
 4. Réaliser κ de Cohen et le nouveau test
- Paramètres : $N = 100, k = 5, rep = 499$

κ de Cohen, $n = 100$ et $k = 5$

p_a	p_d										
	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
0,00	0	0	0	0	0	0	0	0	0	0	0
0,10	0	0	0	0	0	0	0	0	0	0	0
0,20	0,06	0,05	0,05	0,03	0,03	0,03	0,04	0,03	0,04	0,03	0,04
0,30	0,69	0,70	0,72	0,69	0,70	0,71	0,73	0,71	0,71	0,71	0,71
0,40	1	1	1	0,99	1	1	1	1	1	0,99	0,99
0,50	1	1	1	1	1	1	1	1	1	1	1
0,60	1	1	1	1	1	1	1	1	1	1	1
0,70	1	1	1	1	1	1	1	1	1	1	1
0,80	1	1	1	1	1	1	1	1	1	1	1
0,90	1	1	1	1	1	1	1	1	1	1	1

Nouveau test, $n = 100$ et $k = 5$

p_a	p_d									
	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
0,00	0	0	0	0	0	0	0	0	0	0
0,10	0,97	0,88	0,57	0,16	0,01	0	0	0	0	0
0,20	0,95	0,91	0,65	0,23	0,03	0	0	0	0	0
0,30	0,97	0,91	0,68	0,28	0,06	0	0	0	0	0
0,40	0,95	0,92	0,71	0,33	0,09	0	0	0	0	0
0,50	0,94	0,94	0,74	0,44	0,15	0,04	0	0	0	0
0,60	0,94	0,91	0,76	0,47	0,22	0,06	0,02	0	0	0
0,70	0,89	0,89	0,79	0,56	0,34	0,16	0,02	0	0	0
0,80	0,86	0,83	0,80	0,62	0,49	0,27	0,10	0,05	0	0
0,90	0,78	0,82	0,77	0,70	0,61	0,46	0,37	0,26	0,17	0,04

Méthode 2

- On refait la même simulation selon les conditions :
 - $n = 100$ et $k = 3$
 - $n = 100$ et $k = 5$
 - $n = 250$ et $k = 5$
 - $n = 500$ et $k = 5$
 - $n = 1000$ et $k = 2$

Nouveau test, $n = 100$ et $k = 3$

p_a	p_d									
	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
0,00	0	0	0	0	0	0	0	0	0	0
0,10	0,96	0,91	0,78	0,42	0,09	0,02	0	0	0	0
0,20	0,98	0,96	0,81	0,51	0,2	0,05	0	0	0	0
0,30	0,97	0,97	0,84	0,58	0,25	0,07	0,01	0	0	0
0,40	0,98	0,96	0,85	0,62	0,34	0,12	0,02	0	0	0
0,50	0,98	0,96	0,88	0,69	0,42	0,22	0,05	0	0	0
0,60	0,96	0,96	0,9	0,77	0,55	0,27	0,08	0,03	0	0
0,70	0,96	0,94	0,88	0,78	0,64	0,41	0,2	0,06	0,01	0
0,80	0,9	0,9	0,86	0,79	0,71	0,5	0,37	0,18	0,09	0,01
0,90	0,89	0,88	0,84	0,83	0,73	0,71	0,6	0,49	0,34	0,21

Nouveau test, $n = 100$ et $k = 5$

p_a	p_b									
	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
0,00	0	0	0	0	0	0	0	0	0	0
0,10	0,96	0,88	0,57	0,15	0,01	0	0	0	0	0
0,20	0,96	0,92	0,66	0,20	0,02	0	0	0	0	0
0,30	0,96	0,92	0,68	0,30	0,06	0	0	0	0	0
0,40	0,97	0,93	0,76	0,36	0,08	0,01	0	0	0	0
0,50	0,95	0,91	0,76	0,47	0,17	0,02	0	0	0	0
0,60	0,94	0,89	0,83	0,53	0,23	0,04	0	0	0	0
0,70	0,90	0,88	0,76	0,55	0,33	0,12	0,03	0	0	0
0,80	0,86	0,85	0,80	0,60	0,43	0,29	0,11	0,04	0,01	0
0,90	0,81	0,78	0,79	0,66	0,58	0,53	0,33	0,28	0,17	0,10

Nouveau test, $n = 250$ et $k = 5$

p_a	p_b									
	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
0,00	0	0	0	0	0	0	0	0	0	0
0,10	0,95	0,75	0,08	0	0	0	0	0	0	0
0,20	0,96	0,8	0,15	0	0	0	0	0	0	0
0,30	0,97	0,81	0,22	0	0	0	0	0	0	0
0,40	0,95	0,83	0,26	0,02	0	0	0	0	0	0
0,50	0,95	0,83	0,36	0,05	0	0	0	0	0	0
0,60	0,94	0,85	0,42	0,1	0	0	0	0	0	0
0,70	0,92	0,84	0,55	0,2	0,02	0	0	0	0	0
0,80	0,86	0,8	0,57	0,28	0,09	0,01	0	0	0	0
0,90	0,81	0,75	0,7	0,5	0,25	0,13	0,05	0,01	0	0

Nouveau test, $n = 500$ et $k = 5$

p_a	p_b									
	0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90
0,00	0	0	0	0	0	0	0	0	0	0
0,10	0,96	0,46	0	0	0	0	0	0	0	0
0,20	0,96	0,53	0	0	0	0	0	0	0	0
0,30	0,95	0,59	0	0	0	0	0	0	0	0
0,40	0,96	0,64	0,01	0	0	0	0	0	0	0
0,50	0,95	0,68	0,06	0	0	0	0	0	0	0
0,60	0,93	0,7	0,12	0	0	0	0	0	0	0
0,70	0,91	0,73	0,21	0	0	0	0	0	0	0
0,80	0,89	0,73	0,34	0,06	0	0	0	0	0	0
0,90	0,81	0,74	0,51	0,21	0,05	0,01	0	0	0	0

Résultats

- Test sensible à k
 - Plus k est grand, plus le test rejette l'accord
- Test sensible à n
 - Plus n est grand plus le test rejette l'accord

Conclusions

- Réaliser des simulations systématiques sur le nouveau test
- Développer un algorithme pour considérer le pourcentage d'accord
- S'assurer des conditions d'application du test