Université Sidi Mohamed Ben Abdellah
FACULTÉ DES SCIENCES DHAR EL MAHRAZ
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH

DÉPARTEMENT D'INFORMATIQUE

MASTER BIG DATA ANALYTICS & SMART SYSTEMS

MODULE: DATA MINING II

L'intitulé de la présentation:

The Quadratic Weighted Kappa (QWK)

Présentée par :
Otman JAI

Sous la direction de :
Pr. Ismail Berrada
Pr. Afaf Bouhoute

Plan

- Contexte & Introduction
- Aspects mathématiques
- Exemples pratiques avec Sckit-Learn
- Conclusion

Contexte

Compétition Kaggle: Prudential Life Insurance Assessement (PLIA)

Prudential Life Insurance Assessment

Can you make buying life insurance easier?

$30,000
Prize Money

2,610 teams · 4 years ago

Overview | Data | Notebooks | Discussion | Leaderboard | Rules | Team | My Submissions | Late Submission

Overview

Description
Evaluation
Prizes
Timeline
About-Prudential

Submissions are scored based on the quadratic weighted kappa which measures the agreement between two ratings. This metric typically varies from 0 (random agreement) to 1 (complete agreement). In the event that there is less agreement between the raters than expected by chance, this metric may go below 0.

The response variable has 8 possible ratings. Each application is characterized by a tuple *(e,e)*, which corresponds to its scores by *Rater A* (actual risk) and *Rater B* (predicted risk). The quadratic weighted kappa is calculated as follows.

Exemple introductif

- Prenons le cas dans le *domaine médical* où *deux* ou *plusieurs praticiens* **examinant** le *même patient* proposent des *diagnostics différents* ou des décisions thérapeutiques différentes.
- En *l'absence d'une référence*, cette multiplication des avis n'apporte pas la *sécurité* attendue d'un parfait accord diagnostique ou thérapeutique pour le médecin traitant et le patient.
- Il est donc important que *l'accord dans une équipe de travail* ou entre plusieurs équipes soit le meilleur possible pour *garantir la qualité* et la continuité des soins.
- Une ***solution*** consiste ici à réaliser *une séance de «concordance» entre les médecins* pour *estimer leur taux d'accord par le coefficient Kappa* et d'étudier leurs désaccords pour y remédier.

Introduction

Solution: le coefficient (ou test) kappa

- En *statistique*, le test du κ (kappa) **mesure l'accord entre observateurs lors d'un codage qualitatif en catégories**.
- L'article introduisant le κ a pour auteur **Jacob Cohen** – d'où sa désignation de κ de Cohen – et est paru dans le journal Educational and Psychological Measurement en 1960.
- Le κ de Cohen est une mesure d'accord entre *deux évaluateurs seulement*.
- Pour une mesure de l'accord entre *plus de deux évaluateurs*, on utilise le κ de **Fleiss** (1981).

Introduction

Qualitatif vs. Quantitatif

- Le test non paramétrique *Kappa (K) de Cohen* permet de chiffrer l'accord entre deux ou plusieurs observateurs ou techniques lorsque les jugements sont **qualitatifs**.
- Contrairement au *coefficient u de Kendall* par exemple, qui évalue le degré d'accord entre des jugements **quantitatifs**.

Aspects mathématiques

Définition de l'accord

- L'accord entre des jugements est défini comme la conformité de deux ou plusieurs informations qui se rapportent au même objet.
- Cette notion implique l'existence d'une liaison entre les variables, exige des variables de même nature et un appariement des jugements.
- Dans le cas de jugements catégoriels, c'est à dire que la variable aléatoire est discrète et mesurée sur une échelle à **r** niveaux, le taux d'accord ou de «concordance» est estimé par le coefficient Kappa proposé par Cohen en 1960.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Définition mathématique du coefficient Kappa

- L'accord observé entre des jugements qualitatifs ou non, résulte de la somme d'une composante «aléatoire» et d'une composante d'accord «véritable».
- Le coefficient Kappa (K) exprime une différence relative entre la proportion d'accord observée Po et la proportion d'accord aléatoire Pe qui est la valeur espérée sous *l'hypothèse nulle d'indépendance des jugements*, divisée par la quantité disponible au-delà de l'accord aléatoire.
- $K \sim N(K, Var(K))$.
- L'hypothèse nulle H_0 est $K = 0$ contre l'hypothèse alternative $H_1 : K > 0$.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Définition mathématique du coefficient Kappa

- Dans le cas d'une étude *d'accord entre deux observateurs statistiquement indépendants* ayant r modalités de jugement, avec $r \geq 2$, le coefficient Kappa s'écrit :

$$K = \frac{P_O - P_e}{1 - P_e}$$

Avec:

$\boldsymbol{P_o}$: la proportion d'accord observée.

$\boldsymbol{P_e}$: la proportion d'accord aléatoire ou concordance attendue sous l'hypothèse d'indépendance des jugements.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Définition mathématique du coefficient Kappa

- Les deux tableaux suivants présentent la notation utilisée lorsque les données sont présentées dans un *tableau de contingence*.

		Juge A				
	Catégorie	1	2	...	r	Total
Juge B	1	n_{11}	n_{12}	...	n_{1r}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2r}	$n_{2.}$
	.					
	.					
	.					
	r	n_{r1}	n_{r2}	...	n_{rr}	$n_{r.}$
	Total	$n_{.1}$	$n_{.2}$	...	$n_{.r}$	n

Effectifs joints

		Juge A				
	Catégorie	1	2	...	r	Total
Juge B	1	p_{11}	p_{12}	...	p_{1r}	$p_{1.}$
	2	p_{21}	p_{22}	...	p_{2r}	$p_{2.}$
	.					
	.					
	.					
	r	p_{r1}	p_{r2}	...	p_{rr}	$p_{r.}$
	Total	$p_{.1}$	$p_{.2}$	...	$p_{.r}$	1

Proportions jointes

Jugements de deux juges sur une échelle avec r catégories

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Définition mathématique du coefficient Kappa

	Juge A					
	Catégorie	1	2	...	r	Total
Juge B	1	n_{11}	n_{12}	...	n_{1r}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2r}	$n_{2.}$
	.					
	.					
	.					
	r	n_{r1}	n_{r2}	...	n_{rr}	$n_{r.}$
	Total	$n_{.1}$	$n_{.2}$	...	$n_{.r}$	n

	Juge A					
	Catégorie	1	2	...	r	Total
Juge B	1	p_{11}	p_{12}	...	p_{1r}	$p_{1.}$
	2	p_{21}	p_{22}	...	p_{2r}	$p_{2.}$
	.					
	.					
	.					
	r	p_{r1}	p_{r2}	...	p_{rr}	$p_{r.}$
	Total	$p_{.1}$	$p_{.2}$	...	$p_{.r}$	1

On appelle **concordance observée** **P_O**, La proportion des individus classes dans les **cases diagonals** de concordance du tableau de contingence, soit *la somme de ces effectifs diagonaux divisée par la taille de l'échantillon* (**n**):

$$P_O = \sum_{i=1}^{r} p_{ii} = \frac{1}{n}\sum_{i=1}^{r} n_{ii}$$

Avec **r** le *nombre de modalités de jugement*.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Définition mathématique du coefficient Kappa

		Juge A				
	Catégorie	1	2	...	r	Total
Juge	1	n_{11}	n_{12}	...	n_{1r}	$n_{1.}$
B	2	n_{21}	n_{22}	...	n_{2r}	$n_{2.}$
	.					
	.					
	.					
	r	n_{r1}	n_{r2}	...	n_{rr}	$n_{r.}$
	Total	$n_{.1}$	$n_{.2}$	...	$n_{.r}$	n

		Juge A				
	Catégorie	1	2	...	r	Total
Juge	1	p_{11}	p_{12}	...	p_{1r}	$p_{1.}$
B	2	p_{21}	p_{22}	...	p_{2r}	$p_{2.}$
	.					
	.					
	.					
	r	p_{r1}	p_{r2}	...	p_{rr}	$p_{r.}$
	Total	$p_{.1}$	$p_{.2}$	...	$p_{.r}$	1

La **concordance aléatoire** $\mathbf{P_e}$ est égale la somme des produits des effectifs marginaux divisée par le carré de la taille de l'échantillon:

Avec **r** le *nombre de modalités de jugement*.

$$P_e = \sum_{i=1}^{r} p_{i.} p_{.i} = \frac{1}{n^2} \sum_{i=1}^{r} n_{i.} n_{.i}$$

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Remarques

- Le coefficient Kappa est un nombre réel, sans dimension, compris entre -1 et 1.
- L'accord sera d'autant plus élevé que la valeur de Kappa est proche de 1.
- L'accord maximal est atteint ($K = 1$) lorsque $Po = 1$ et $Pe = 0{,}5$.
- Lorsqu'il y a indépendance des jugements, le coefficient Kappa est égal à zéro ($Po = Pe$).
- Dans le cas d'un désaccord total entre les juges, le coefficient Kappa prend la valeur -1 avec $Po = 0$ et $Pe = 0{,}5$.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Interprétation du coefficient Kappa: Degré d’accord

Landis et Koch ont proposé un classement de l’accord en fonction de la valeur de Kappa présenté dans le tableau suivant:

Accord	*Kappa*
Excellent	$\geq 0{,}81$
Bon	0,80 - 0,61
Modéré	0,60 - 0,41
Médiocre	0,40 - 0,21
Mauvais	0,20 - 0,0
Très mauvais	$< 0{,}0$

Les limites de ce classement sont arbitraires et peuvent varier selon l’étude réalisée. Dans tous les cas, le classement devra être défini avec des experts avant la réalisation de l’étude.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Exemple de calcul du Kappa

On souhaite *évaluer le degré d'accord* entre les *réponses positives et négatives* fournies par deux ***tests biologiques A et B*** appliqués aux *mêmes échantillons sériques*. L'étude porte sur *200 échantillons* et les résultats sont présentés dans le tableau suivant:

Résultats des tests A et B appliqués aux mêmes échantillons

		Résultat du test A		
	Réponses	+	-	Total
Résultat du test B	+	**72**	**16**	88
	-	**25**	**87**	112
	Total	97	103	200

Les deux tests sont en accord pour 159 échantillons avec 72 réponses positives concordantes et 87 réponses négatives concordantes.

La proportion d'accord observé et la proportion d'accord aléatoire sont :

$$P_O = \frac{72+87}{200} = 0.795$$

$$P_e = \frac{(88\times 97)+(112\times 103)}{200^2} = 0.5018$$

$$K = \frac{P_O - P_e}{1-P_e} = \frac{0.795-0.5018}{1-0.5018} \approx 0.59$$

Cette valeur indique un *accord modéré* entre les deux tests.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Exemple de calcul du Kappa: Accord maximal

La valeur maximale de Kappa (K_m) est définie par la relation suivante:

$$K_m = \frac{P_m - P_e}{1 - P_e}$$

P_m est la proportion de l'accord maximale définie par:

$$P_m = \sum_{i=1}^{r} \inf\left(p_{i.}, p_{.i}\right)$$

$p_{i.}$ et **$p_{.i}$** sont les effectifs marginaux.

Résultats des tests A et B appliqués aux mêmes échantillons

		Résultat du test A		
	Réponses	+	-	Total
Résultat du test B	+	**72**	**16**	88
	-	**25**	**87**	112
	Total	97	103	200

$$P_m = \frac{\inf\left(88,97\right) + \inf\left(112,103\right)}{200} = \frac{88 + 103}{200} = 0.955$$

$$K_m = \frac{0.955 - 0.5018}{1 - 0.5018} \approx 0.91$$

$$\frac{K}{K_m} = \frac{0.59}{0.91} \times 100 \approx 65\%$$

En conclusion, l'accord obtenu entre les deux tests biologiques correspond à 65% de l'accord maximal qu'il pourrait atteindre.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Le coefficient Kappa pondéré (K_w): Introduction

- Certaines discordances entre les juges sont plus graves que d'autres.
- Cohen propose de donner à chacune des cases du tableau de contingence, un **poids w_{ij}** fixé a priori qui *reflète l'importance que l'on attribue au désaccord*.
- On utilise le plus souvent des ***poids de concordance*** plutôt que des poids de discordance.
- Ceci peuvent varier de 1 pour les cases diagonales à 0 pour les cases qui correspondent au plus grand désaccord en considérant que l'échelle des *catégories de jugements est ordonnée*.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Le coefficient Kappa pondéré (K_w): Définition

- Au tableau de contingence **r**x**r** représentant les résultats d'une étude d'accord, nous associons une matrice de poids **r**x**r** notée **W** définissant l'importance de chaque désaccord.

		Juge A			
	Catégorie	1	2	...	r
Juge B	1	w_{11}	w_{12}	...	w_{1r}
	2	w_{21}	w_{22}	...	w_{2r}
	.				
	.				
	.				
	r	w_{r1}	w_{r2}	...	w_{rr}

Matrice de poids W associée à un tableau de contingence r x r

Régle de remplissage de la matrice des poids

1 pour les cases diagonales à **0** pour les cases qui correspondent au plus grand désaccord en *considérant que l'échelle des catégories de jugements est ordonnée.*

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Le coefficient Kappa pondéré (K_w): Systèmes de Pondération

- Le choix des poids du ***Kappa pondéré K_w*** peut être réalisé selon deux types de systèmes de pondération:
  - Système de *pondération linéaire*:

$$w_{ij} = 1 - \frac{|i-j|}{r-1}$$

  - Système de *pondération quadratique*:

$$w_{ij} = 1 - \frac{(i-j)^2}{(r-1)^2}$$

Avec:

- i: i$^{\text{ème}}$ colonne de la matrice des poids
- j: j$^{\text{ème}}$ ligne de la matrice des poids
- r: le nombre de modalités de jugement
- w_{ij} : le poids de la case (i,j) du tableau de contingence

Remarque importante:
La matrice des poids sera choisie <u>symétrique</u> dans le cas d'une étude de <u>reproductibilité</u> et pour d'autres types d'étude elle peut être choisie <u>asymétrique</u> si l'on désire <u>souligner une dissymétrie entre les juges</u>.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Le coefficient Kappa pondéré (K_w): Expression Mathématique

- La concordance observée $\mathbf{P_{o(w)}}$ du kappa pondéré en fonction de la matrice des poids de concordance est définie par :

$$P_{O(w)} = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} p_{ij}$$

- Et la concordance aléatoire $\mathbf{P_{e(w)}}$ est:

$$P_{e(w)} = \sum_{i=1}^{r}\sum_{j=1}^{r} w_{ij} p_{i.} p_{.j}$$

- Le ***Kappa pondéré*** est donné par:

$$K_w = \frac{P_{O(w)} - P_{e(w)}}{1 - P_{e(w)}}$$

Avec:

$\mathbf{p_{ij}} = \mathbf{n_{ij}}/n$

$\mathbf{p_{i.}} = \mathbf{n_{i.}}/n$

$\mathbf{p_{.j}} = \mathbf{n_{.j}}/n$

n étant le nombre total d'observations.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Signification Statistique: Erreur standard de la concordance aléatoire

- Pour ***tester l'hypothèse*** nulle que *les jugements sont indépendants* (H_0 : K = 0), c'est à dire que la seule liaison entre les jugements est due au hasard, ***Fleiss***, ***Cohen*** et ***Everitt***, ont montré que l'*erreur standard de la concordance aléatoire* S_{K0} est estimée par :

$$S_{K_0} = \frac{1}{(1-P_e)\sqrt{n}}\sqrt{P_e + P_e^2 - C}$$

$$Avec: \ C = \sum_{i=1}^{r} p_{i.}p_{.i}\left(p_{i.} + p_{.i}\right)$$

- Et pour le ***Kappa pondéré:***

$$S_{K_{w0}} = \frac{1}{\left(1-P_{e(w)}\right)\sqrt{n}}\sqrt{P_{e(w)} + P_{e(w)}^2 - C}$$

$$Avec: \ C = \sum_{i=1}^{r}\sum_{j=1}^{r} p_{i.}p_{.i}\left[w_{ij} - \left(\bar{w}_{i.} + \bar{w}_{.j}\right)\right]^2$$

$$Et: \ \bar{w}_{i.} = \sum_{j=1}^{r} w_{ij}p_{.j} \ \text{ et } \ \bar{w}_{.j} = \sum_{i=1}^{r} w_{ij}p_{i.}$$

La variable aléatoire centrée réduite Z du coefficient *K*:

$$Z = \frac{K-0}{S_{K_0}}$$

Sous l'hypothèse nulle H_0 : «indépendance des jugements» (d'où $K = 0$) contre l'hypothèse alternative H_1 : $K > 0$, Z suit une **loi normale centrée réduite**.

$Si\ Z \succ Z_{1-\alpha}$, on rejette H_0 pour un risque α unilatéral.

Ceci est valable pour une taille de l'échantillon **(n)** > **2r²** avec **r** le nombre de modalités.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Signification Statistique : Erreur standard du coefficient Kappa

- Pour ***tester l'hypothèse*** nulle que *les classements effectués par les deux juges sont indépendants* (H_0 : $p_{ij} = p_{i.}p_{.j}$ d'où Po = Pe et K = 0), ***Fleiss***, ***Cohen*** *et* ***Everitt***, ont montré que l'*erreur standard de Kappa* S_K est estimée par :

$$S_K = \frac{\sqrt{A+B-C}}{(1-P_e)\sqrt{n}}$$

$$Avec: \; C = \left[K - P_e(1-K)\right]^2$$

$$A = \sum_{i=1}^{r} p_{ii}\left[1-(p_{i.}+p_{.i})(1-K)\right]^2$$

$$B = (1-K)^2 \sum_{i=1}^{r}\sum_{j=1, i\neq j}^{r} p_{ij}\left(p_{.i}+p_{j.}\right)^2$$

- Et pour le ***Kappa pondéré:***

$$S_{K_w} = \frac{1}{(1-P_{e(w)})\sqrt{n}}\sqrt{A-B}$$

$$Avec: \; A = \sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij}\left[w_{ij} - (\overline{w}_{i.} + \overline{w}_{.j})(1-K_w)\right]^2$$

$$Et: \; B = \left[K_w - P_{e(w)}(1-K_w)\right]^2$$

La variable aléatoire centrée réduite Z du coefficient *K*:

$$Z = \frac{K-0}{S_{K_0}}$$

Sous l'hypothèse nulle H_0 : «indépendance des classements» (d'où $K = 0$) contre l'hypothèse alternative H_1 : $K > 0$, Z suit une **loi normale centrée réduite** pour un échantillon de taille **n ≥ 3r²** (comparer 2 K) et **n ≥ 16r²** pour déterminer l'intervalle de confiance du Kappa.

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Exemple Pratique: Estimation de l'accord entre deux neurologues

- Ce type d'étude nous permet d'estimer l'accord entre deux juges avec ou sans matrice de poids de concordance.
- Cette exemple porte sur l'évaluation de l'accord entre deux médecins dans le diagnostic clinique de la *sclérose en plaques* (التصلب اللويحي).
- Les neurologues devaient classer 149 patients dans l'une des <u>*quatre catégories*</u> de diagnostic de la sclérose en plaques :

1. **certaine ;**
2. **probable ;**
3. **possible (une chance sur deux) ;**
4. **douteuse, incertaine ou pas de sclérose en plaques.**

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Exemple Pratique: Estimation de l'accord entre deux neurologues

Classification des patients par les deux neurologues

		Neurologue n°1				
	Diagnostic	1	2	3	4	Total
Neurologue n°2	1	**38**	**5**	**0**	**1**	44
	2	**33**	**11**	**3**	**0**	47
	3	**10**	**14**	**5**	**6**	35
	4	**3**	**7**	**3**	**10**	23
	Total	84	37	11	17	**149**

- Le Kappa général est médiocre selon la classification proposée par **Landis** et **Koch**, mais il est significativement différent de zéro c'est à dire que les jugements ne sont pas indépendants.
- Les accords catégoriels sont tous significatifs au risque de 5%, sauf l'accord de la catégorie 2 qui peut être expliqué par le hasard.

Résultat du Kappa non pondéré entre les deux neurologues

Catégorie	Po	Pe	Kappa	S_{K0}	Z	p	S_K
1	0,651	0,474	0,337	0,071	4,762	$< 1{,}0E^{-5}$	0,065
2	0,584	0,593	-0,022	0,081	-0,273	0,392	0,080
3	0,758	0,726	0,118	0,067	1,780	0,038	0,081
4	0,866	0,767	0,424	0,081	5,243	$< 1{,}0E^{-6}$	0,106
Général	**0,430**	**0,280**	**0,208**	**0,046**	**4,544**	$< \mathbf{1{,}0E^{-5}}$	**0,051**

Aspects mathématiques

Accord entre 2 juges: Kappa de Cohen

Exemple Pratique: Estimation de l’accord entre deux neurologues

Par cette matrice, on transforme les désaccords de jugements en des concordances, entre les catégories 1 et 2 et entre les catégories 3 et 4.

Matrice des poids de concordance entre les deux neurologues

		Neurologue n°1			
	Diagnostic	1	2	3	4
Neurologue n°2	1	1,0	1,0	0,0	0,0
	2	1,0	1,0	0,0	0,0
	3	0,0	0,0	1,0	1,0
	4	0,0	0,0	1,0	1,0

Résultat global du Kappa pondéré entre les deux neurologues

Po	*Pe*	Kappa	S_{K0}	*Z*	*p*	S_K
0,745	0,569	**0,408**	0,073	5,616	$< 1{,}0E^{-6}$	0,072

L’accord pondéré global des neurologues est modéré mais il est significativement supérieur à la valeur précédente du Kappa sans pondération au risque d’erreur de 5% (test bilatéral).

Exemples pratiques

Implémentation de la mesure Quadratic Weighted Kappa (QWK)

Quadratic Kappa Metric explained in 5 simple steps

Python notebook using data from PetFinder.my Adoption Prediction · 20,455 views · 1y ago · beginner, tutorial, statistics

270

Edit: Quadratic Kappa Metric is the same as cohen kappa metric in Sci-kit learn @ sklearn.metrics.cohen_kappa_score when weights are set to 'Quadratic'. Thanks to Johannes for figuring that out.

https://www.kaggle.com/aroraaman/quadratic-kappa-metric-explained-in-5-simple-steps

Conclusion

À partir de l'ensemble de ces résultats on peut dire que:

- **Le test statistique Kappa de Cohen permet d'estimer, de déceler et de quantifier les désaccords pour les corriger.**

- **Ce test est un instrument précieux pour le contrôle de la qualité des techniques et des soins mais son interprétation exige une bonne connaissance de ces limites.**

- **Le domaine d'application du coefficient Kappa est large, et des études d'accord, aussi bien entre des observateurs ou des examens, devraient être spontanément ou régulièrement réalisées dans le cadre de l'assurance qualité des soins et des techniques.**

Merci pour votre attention