

Échelle sociale dérivée de comparaisons entre professions

Tiaray RAZAFINDRANOVONA

Insee, division Emploi

14 juin 2018



Problématique : hiérarchie des professions

- Existence d'une hiérarchie sous-jacente entre professions ?
- Classement selon indicateurs socio-économiques :
lien avec réalité concrète, opinions, rapports de force ?
- Travaux sur l'évaluation sociale des professions :
Chambaz *et al.* (1998) à partir d'une enquête expérimentale
- Enquête FQP 2014-2015 :
questions de comparaisons entre professions
- Application de modèles de comparaison par paires



Plan

- 1 Classement dérivé de comparaisons deux à deux
- 2 Application des modèles de Bradley-Terry à FQP 2014-2015
- 3 Discussion



Plan

- 1 Classement dérivé de comparaisons deux à deux
- 2 Application des modèles de Bradley-Terry à FQP 2014-2015
- 3 Discussion



Données de comparaisons par paires

- Données collectées relatives aux préférences prennent souvent la forme de comparaisons par paires
- Plus facile, plus pratique d'effectuer des comparaisons locales
- Confrontations entre participants à une compétition sportive
- Classement des objets comparés ?



Le modèle de Bradley-Terry (1952)

- Modèle de Bradley-Terry (BT) : le plus populaire des modèles utilisés sur des données de comparaisons par paires en vue d'inférer un classement
- Principe : estimer pour chacun des objets r , un paramètre de score ψ_r reflétant sa force, son mérite, son prestige ou sa compétence
- Hypothèse sous-jacente : **existence d'une échelle unidimensionnelle du mérite**, l'estimation des paramètres ψ_r fournit alors un classement de l'ensemble des objets



Relation fondamentale

Considérant une relation de domination \succ , Bradley et Terry suggèrent que la probabilité qu'un objet r domine un objet s s'exprime sous la forme :

$$P(r \succ s) = \frac{\psi_r}{\psi_r + \psi_s} \quad (1)$$

Dans sa version originelle, le modèle BT ne s'applique qu'à des données de confrontations à issues binaires.



Historique et exemples d'applications

Ces principes se retrouvent dans des travaux du mathématicien Zermelo (1929) : son objectif était d'estimer le niveau des joueurs d'échecs à l'issue d'un tournoi sachant que tous ne s'étaient pas affrontés le même nombre de fois.

Exemples d'application

- Marketing, psychologie expérimentale : données recueillies sous la forme de **jugements de personnes**
- Sport : données de **confrontations entre participants**
- Autres exemples : génétique, influence de revues académiques



Estimation (1)

Considérant un ensemble de M objets et la modélisation (1), la log-vraisemblance s'écrit :

$$l(\psi) = \sum_{r=1}^M \sum_{s=1}^M [w_{rs} \ln(\psi_r) - w_{rs} \ln(\psi_r + \psi_s)]$$

avec w_{rs} le nombre de fois où l'objet r domine l'objet s , avec la convention $w_{rr} = 0$.

L'estimation est réalisée par **maximum de vraisemblance**.



Estimation (2)

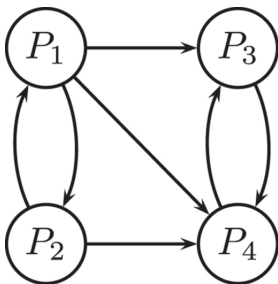
CNS d'existence et d'unicité du MV

Pour toute partition d'objets dans deux ensembles non vides, l'un des objets du second ensemble domine l'un des objets du premier ensemble au moins une fois.



Estimation (3)

CNS : graphe orienté associé aux données est **fortement connexe**



Graphique: Exemple de graphe **non** fortement connexe

Estimation (4)

Une reparamétrisation simple ramène l'estimation du modèle de Bradley-Terry à celle d'un **modèle logistique** ce qui a aidé historiquement à la popularisation de ces modèles.

$$\psi_r = e^{\gamma_r} \quad \psi_s = e^{\gamma_s}$$
$$\text{logit } P(r \succ s) = \ln \frac{P(r \succ s)}{1 - P(r \succ s)} = \gamma_r - \gamma_s$$

Classe plus large de modèles tels que :

$$P(r \succ s) = F(\mu_r - \mu_s)$$

où $F(-\infty) = 0$, $F(+\infty) = 1$ et $F(-x) = 1 - F(x)$.

F fonction de répartition de la loi logistique : modèle BT

F fonction de répartition de la loi normale : modèle de Thurstone



Principales extensions au modèle BT

- Cas d'égalités ou issues plus nombreuses et ordonnées
- Effet d'ordre (avantage du terrain en sport)
- Covariables dans la modélisation
 - variables spécifiques aux *objets*
 - variables spécifiques aux *sujets*
 - variables relatives aux *objets* et aux *sujets*

objets : les entités comparées qu'on souhaite classer

sujets : les entités qui confrontent les objets

Packages R utilisés : BradleyTerry2 et BTLasso



Ex : football, rencontres internationales (hors amicaux)



1998-2018			2014-2018		
Rang	Pays	Coefficient BT	Rang	Pays	Coefficient BT
1	Spain	3,717	1	Germany	5,173
2	Germany	3,687	2	Brazil	4,614
3	Brazil	3,550	3	Argentina	4,547
4	Argentina	3,534	4	Portugal	4,502
5	Netherlands	3,396	5	France	4,371
6	France	3,290	6	Poland	4,308
7	Italy	3,247	7	Chile	4,294
8	Portugal	3,203	8	Colombia	4,115
9	England	3,129	9	Spain	4,106
10	Mexico	2,855	10	Peru	3,969
11	Croatia	2,838	11	Belgium	3,883
12	Colombia	2,834	12	England	3,882
13	Czech.Republic	2,789	13	Switzerland	3,853
14	Chile	2,730	14	Uruguay	3,752
15	USA	2,719	15	Iceland	3,654
16	Sweden	2,710	16	Italy	3,648
17	Uruguay	2,685	17	Mexico	3,526
18	Denmark	2,590	18	Paraguay	3,465
19	Switzerland	2,570	19	Netherlands	3,391
20	Paraguay	2,527	20	Wales	3,317

Tableau: Classements fondés sur des estimations de modèles BT



Plan

- 1 Classement dérivé de comparaisons deux à deux
- 2 Application des modèles de Bradley-Terry à FQP 2014-2015
- 3 Discussion



Enquête FQP 2014-2015

- 7ème édition d'une enquête réalisée par l'Insee en 1964, 1970, 1977, 1985, 1993 et 2003
- une des principales sources d'information sur la mobilité professionnelle, la mobilité sociale et les relations entre la formation initiale et professionnelle, l'emploi et les salaires
- environ 27 000 répondants
- questions de comparaisons de professions (père, mère, N-5)



Question de comparaison entre professions

QF - 11. Revenons à votre profession principale actuelle. Si vous la comparez à la profession de votre père (ou tuteur, ou père adoptif) que nous venons de décrire, diriez-vous que son niveau ou son statut est... **Ne pas lire la modalité « Pas comparable »**

1. Bien plus élevé que celui de votre père (ou tuteur, ou père adoptif)
2. Plus élevé
3. A peu près le même
4. Plus bas
5. Bien plus bas
6. Ne pas lire : Pas comparable

Question mobilisée pour confronter la mobilité sociale intergénérationnelle observée au ressenti des personnes (Razafindranovona, 2017).



Les catégories de professions

Professions évoquées dans la question ont été décrites lors du questionnement et codées (en direct ou *a posteriori*). Trois niveaux de regroupements sont étudiés ici :

- le groupe social correspondant au niveau agrégé de la nomenclature PCS 2003 en 6 postes : agriculteurs exploitants, artisans, commerçants et chefs d'entreprise, cadres et professions intellectuelles supérieures, professions intermédiaires, employés et ouvriers ;
- un regroupement *ad hoc* reprenant le groupe social en distinguant au sein des employés, employés qualifiés et employés non qualifiés et au sein des ouvriers, ouvriers qualifiés et ouvriers non qualifiés ;
- le niveau détaillé de la catégorie socioprofessionnelle.



Groupe social

	Coefficient BT	Erreur standard
3. Cadres et professions intellectuelles supérieures	2,988	0,061
4. Professions Intermédiaires	1,588	0,042
2. Artisans, commerçants et chefs d'entreprise	1,418	0,053
5. Employés	0,425	0,036
6. Ouvriers	0,000	
1. Agriculteurs exploitants	-0,542	0,078

Tableau: Modèle originel de Bradley-Terry, comparaison avec père, groupe social



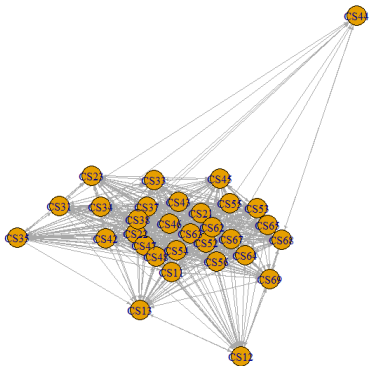
Regroupement *ad hoc*

	Coefficient BT	Erreur standard
3. Cadres et professions intellectuelles supérieures	3,659	0,074
4. Professions Intermédiaires	2,246	0,060
2. Artisans, commerçants et chefs d'entreprise	2,056	0,068
5a. Employés qualifiés	1,357	0,060
6a. Ouvriers qualifiés	0,845	0,054
5b. Employés non qualifiés	0,466	0,066
1. Agriculteurs exploitants	0,009	0,087
6b. Ouvriers non qualifiés	0,000	

Tableau: Modèle original de Bradley-Terry, comparaison avec père, regroupement *ad hoc*



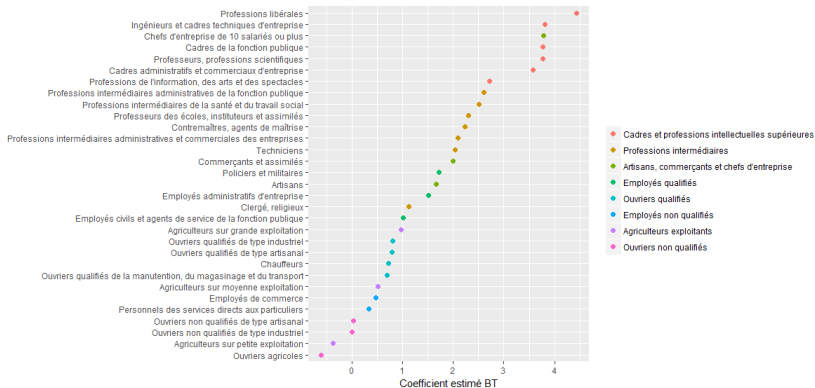
Graphe orienté



Graphique: Graphe orienté, comparaison avec père, CS détaillée



CS détaillée



Graphique: Modèle original de Bradley-Terry, comparaison avec père, CS détaillée



Principaux enseignements (robustes aux différents périmètres et spécifications)

- Hiérarchie assez claire sur les groupes sociaux de salariés, conforme avec le principe hiérarchique sous-jacent des groupes sociaux de salariés de la nomenclature PCS
- Employés Q et ouvriers Q à un niveau sensiblement supérieur à celui des employés NQ et des ouvriers NQ, pertinence de ce clivage pour décrire le bas de l'échelle sociale
- Homogénéité des groupes sociaux de salariés : pas ou peu de chevauchements dans les positions des CS détaillées selon leur catégorie *ad hoc* d'appartenance
- Quelques professions se distinguent néanmoins : chez les cadres, professions de l'information, des arts et des spectacles
- Forte hétérogénéité dans les groupes sociaux d'indépendants : chefs d'entreprise dans le haut de la hiérarchie, artisans et commerçants à un niveau plus intermédiaire



Confrontation avec d'autres logiques de classement

- Capital culturel et/ou économique
 - en haut de la hiérarchie BT : professions cumulant capital culturel et économique (professions libérales), professions à gros capital culturel (professeurs et professions scientifiques), professions à gros capital économique (chefs d'entreprise)
 - plus que le patrimoine ou les diplômes, ce sont les revenus qui expliquent les positions en bas de l'échelle sociale
- Travaux de Chambaz *et al.* (1998) sur l'évaluation sociale des professions
 - à partir d'une enquête expérimentale, 3 000 répondants, qui notent des professions
 - mêmes tendances globales : chefs d'entreprise un peu plus bas que dans les modèles BT, agriculteurs un peu plus haut



Plan

- 1 Classement dérivé de comparaisons deux à deux
- 2 Application des modèles de Bradley-Terry à FQP 2014-2015
- 3 Discussion



Discussion

- Pas un indicateur unique et définitif : mais éléments pour appréhender la hiérarchie sociale sous-jacente
- Cadre diffère un peu de l'utilisation habituelle en marketing : lien fort entre objets et sujets
- Comparaisons extrapolées sur des codifications de professions
- Représentativité : dans quelle mesure infère-t-on la hiérarchie sous-jacente de la société française à partir de telles données ?
- Utilisation de covariables : que peut-on souhaiter contrôler pour obtenir un tel classement ?

Au final, une échelle sociale plutôt conforme aux grandes logiques de regroupements et de catégorisations savantes de l'espace social malgré quelques dissonances ; ces logiques semblent bien, dans une certaine mesure, intériorisées lorsque les personnes effectuent leurs comparaisons !



Bibliographie

Bradley, R. A., Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324-345.

Chambaz, C., Maurin, É., Torelli, C. (1998). L'évaluation sociale des professions en France: Construction et analyse d'une échelle des professions. *Revue française de sociologie*, 177-226.

Razafindranovona, T. (2017). Malgré la progression de l'emploi qualifié, un quart des personnes se sentent socialement déclassées par rapport à leur père. *Insee Première*, 1659.

Zermelo, E. (1929). Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29(1), 436-460.



Merci de votre attention!

