Evaluation de la capacité de mesures d'inégalité à détecter des changements dans une distribution de revenus

Matti Langel Yves Tillé

Institut de statistique
Université de Neuchâtel, Suisse

Journées de méthodologie statistique, 2009

Plan

Introduction

- Une multitude d'indices synthétiques permettent de mesurer l'inégalité dans une distribution de revenus.
- Objectifs :
  - Parvenir à détecter et mesurer un (petit) changement dans le niveau d'inégalité.
  - Etudier quelle mesure permet de détecter quel type de changement.
  - Evaluer la performance des indicateurs de Laeken (*QSR* et Gini).
- Méthode : comparer par simulations le niveau d'inégalité (estimé sur un échantillon) dans une distribution avant et après perturbation.

Notation

- Population finie U, de taille N.
- y_i : le revenu de l'observation i.
- Distribution ordonnée : $\text{rang}(y_i) = i$.
- Q_p : le quantile d'ordre p.
- $Y, \overline{Y}$ et $Var(y)$ représentent respectivement le revenu total, le revenu moyen et la variance dans la population.
- Quantile shares :

$$S_p^+ = \sum_{y_i \geq Q_p} y_i, \quad S_p^- = \sum_{y_i \leq Q_p} y_i.$$

Indices I : Indice de Gini, Coefficient de variation

Indice de Gini

$$Gini = \frac{2\sum_{i=1}^{N} iy_i}{N\sum_{i=1}^{N} y_i} - \frac{1}{N} - 1.$$

- L'indice d'inégalité le plus utilisé dans la littérature.
- L'un des deux indicateurs d'inégalité de Laeken (Eurostat).

Coefficient de variation

$$CV = \frac{\sqrt{Var(y)}}{\overline{Y}}.$$

Indices II : Indices d'Atkinson

- **Indices d'Atkinson** :
  - Basé sur la SWF (fonction de "bien-être social").
  - α : paramètre d'"aversion aux inégalités"
    (forte aversion lorsque α est proche de 1).
  - Forme générale :

$$A_\alpha = \begin{cases} 1 - \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i}{\bar{\bar{Y}}} \right)^{1-\alpha} \right]^{\frac{1}{1-\alpha}}, & \text{si} \quad 0 \leq \alpha < 1, \\ 1 - \prod_{i=1}^{N} \left(\frac{y_i}{\bar{\bar{Y}}} \right)^{\frac{1}{N}}, & \text{si} \quad \alpha = 1. \end{cases}$$

Indices III : Indices d'Atkinson

Deux indices d'Atkinson sont traités dans nos simulations :

Atkinson ($\alpha = 0.5$)

$$A_{1/2} = 1 - \frac{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{y_i}\right)^2}{\overline{Y}}.$$

Atkinson ($\alpha = 1$)

$$A_1 = 1 - \prod_{i=1}^{N}\left(\frac{y_i}{\overline{Y}}\right)^{\frac{1}{N}} = 1 - \frac{G}{\overline{Y}}.$$

avec G, la moyenne géometrique : $G = \left(\prod_{i=1}^{N} y_i\right)^{\frac{1}{N}}$.

Indices IV : Quantile Share Ratios

- famille d'indices basés sur les *quantile shares*.
- Le *Quintile* Share Ratio (*QSR*) est le deuxième indicateur de Laeken.

Quintiles (*QSR*), déciles (*DSR*), médiane (*MSR*)

$$QSR = \frac{S^{+}_{80}}{S^{-}_{20}}, \qquad DSR = \frac{S^{+}_{90}}{S^{-}_{10}}, \qquad MSR = \frac{S^{+}_{50}}{S^{-}_{50}}.$$

- Une nouvelle mesure :

Average Share Ratio (*ASR*)

$$ASR = N \left(\sum_{j=1}^{N} \frac{\sum_{i=1}^{j} y_i}{\sum_{i=N-j+1}^{N} y_i} \right)^{-1}.$$

Indices V : Indice de Zenga

- proposé par Michele Zenga (2007).
- basé sur des ratios de moyennes arithmétiques.

Indice de Zenga

$$Z = 1 - \frac{1}{N}\sum_{j=1}^{N}\frac{\overline{Y}_j^-}{\overline{Y}_j^+}.$$

$$\text{avec} \quad \overline{Y}_j^- = \frac{1}{j}\sum_{i=1}^{j} y_i \quad \text{et} \quad \overline{Y}_j^+ = \frac{1}{N-j+1}\sum_{i=j}^{N} y_i.$$

Indices VI : Entropie généralisée

- Forme générale : $GE_\alpha = \frac{\frac{1}{N}\sum_{i=1}^{N}(\frac{y_i}{\bar{Y}})^\alpha - 1}{\alpha(\alpha-1)}.$
- Quand α tend vers 1 :

Indice de Theil (T)

$$GE_1 = T = \sum_{i=1}^{N} \frac{y_i}{Y} \log \frac{Ny_i}{Y}.$$

- Deux autres cas utilisés :

Entropie généralisée avec $\alpha = 0.5$ ($GE_{1/2}$) et $\alpha = 2$ (GE_2)

$$GE_{1/2} = 4\left(1 - \sum_{i=1}^{N}\sqrt{\frac{y_i}{NY}}\right) \quad \text{et} \quad GE_2 = \frac{1}{2N}\sum_{i=1}^{N}\frac{y_i^2}{\bar{Y}^2} - \frac{1}{2}.$$

Indices VII : Divergence de Renyi

- Forme générale : $R_\alpha = \frac{1}{\alpha - 1} \log N^{\alpha-1} \sum_{i=1}^{N} \left(\frac{y_i}{Y}\right)^{\alpha}$.
- Deux cas utilisés :

Divergence de Renyi avec $\alpha = 0.5$ ($R_{1/2}$) et $\alpha = 2$ (R_2)

$$R_{1/2} = -2 \log \sum_{i=1}^{N} \sqrt{\frac{y_i}{NY}} \quad \text{et} \quad R_2 = \log N \sum_{i=1}^{N} \frac{y_i^2}{Y^2}.$$

- Note : Lorsque α tend vers 1, on retrouve l'indice de Theil.

Indices VIII : Rapports Interquantiles

- Simples rapports de quantiles ($\neq$ *quantile shares*).

Rapports de quartiles (*IQR*), de déciles (*IDR*)

$$IQR = \frac{Q_{75}}{Q_{25}} \quad \text{et} \quad IDR = \frac{Q_{90}}{Q_{10}}.$$

- Une nouvelle mesure :

Mean Quantile Ratio (*MQR*)

$$MQR = \frac{N}{2\sum_{i=1}^{N/2} \frac{y_i}{y_{N-i+1}}}.$$

Simulations : les données

- Distribution initiale (distribution *A*) :
  - Revenu imposable annuel des ménages du canton de Neuchâtel.
  - 88'011 contribuables au revenu non nul.
- Distribution modifiée (distribution *B*) :
  - Obtenue en **appliquant une perturbation** à la distribution *A*.
- Simulations réalisées avec **11** perturbations différentes.

Perturbations en haut de la distribution

1. **Plafonnement des hauts revenus.**
   $\Rightarrow$ Nivellement de tous les revenus supérieurs à Q_{90} au revenu de Q_{90}.
2. **Doublement des très hauts revenus.**
   $\Rightarrow$ Doublement du revenu des 5% les plus riches.
3. **Réductions des très hauts revenus.**
   $\Rightarrow$ Réduction de moitié du revenu des 5% les plus riches.

Perturbations sur toute la distribution

4. **Translation.**
   $\Rightarrow$ Ajout d'un revenu fixe de CHF 10'000 à toutes les observations.

5. **Reduction de la dispersion de la distribution.**
   $\Rightarrow y_{i_B} = 0.9 y_{i_A} + 0.1 \overline{Y}_A.$

6. **Revenus après imposition.**
   $\Rightarrow$ Le montant réel de la taxation est fourni par la base de données.

Perturbations au bas de la distribution

7. **Revenu minimum garanti.**
   $\Rightarrow$ Pour cette simulation, tous les plus bas revenus ont été nivellés à un revenu égale à 30% du revenu médian.
8. **Augmentation des plus bas revenus.**
   $\Rightarrow$ Le revenu des 5% les plus pauvres est multiplié par 10.
9. **Diminution des plus bas revenus.**
   $\Rightarrow$ Le revenu des 5% les plus pauvres est divisé par 10.

Perturbation au centre, évaluation de la robustesse

10. **Reduction de la dispersion au centre de la distribution.**
$$\Rightarrow y_{i_B} = \begin{cases} 0.1 y_{i_A} + 0.9\overline{Y}_A, & \text{si } Q_{30} \leq y_{i_A} \leq Q_{70} \\ y_{i_A}, & \text{sinon.} \end{cases}$$

11. **Evaluation de la robustesse.**
$\Rightarrow$ Sensibilité à une valeur extrême : un revenu sélectionné aléatoirement a été multiplié par dix mille.
Note : Cette perturbation est appliquée directement aux échantillons, non à la population.

Simulations : la méthode

Pour chacune des 11 perturbations :

- Séléction d'un échantillon S_A (plan simple sans remise, $n = 1000$) issu de la distribution A et de S_B (idem) issu de B.
- Estimation de chaque indice d'inégalité I sur S_A et sur S_B.
- 10'000 réplications.
- Calcul de $\overline{\widehat{I}_A}$, $\overline{\widehat{I}_B}$, $Var(\widehat{I}_A)$ et $Var(\widehat{I}_B)$, pour chaque indice d'inégalité sur les 10'000 simulations.
- Calcul de la statistique z :

$$z = \left| \frac{\overline{\widehat{I}_B} - \overline{\widehat{I}_A}}{\sqrt{Var(\widehat{I}_A) + Var(\widehat{I}_B)}} \right| .$$

$\Rightarrow$ une forte valeur de z indique une forte capacité à déceler le changement du niveau d'inégalités entre les deux distributions.

Simulations : Tableau des résultats (statistique z)

	haut de la distribution			toute la distribution			bas de la distribution			centre	robust.
	1	2	3	4	5	6	7	8	9	10	11
Gini	3.52	2.86	2.53	2.67	1.76	0.78	0.75	1.01	0.14	1.49	5.52
CV	1.01	0.78	0.65	0.28	0.20	0.12	0.04	0.07	0.01	0.12	5.40
$A_{1/2}$	2.37	2.28	1.72	2.19	1.54	0.55	0.99	1.04	0.51	0.50	5.31
A_1	2.29	2.47	1.77	4.24	**3.14**	0.78	2.58	2.43	**2.34**	0.41	5.48
QSR	2.21	2.12	1.81	4.10	2.92	0.97	2.29	2.34	0.43	0.01	0.96
DSR	2.11	1.93	1.50	3.79	2.96	0.72	**3.49**	**2.79**	0.92	0.01	0.91
MSR	2.51	2.33	2.12	3.52	2.37	1.02	0.90	1.45	0.16	3.13	0.99
ASR	3.03	2.49	2.30	3.16	2.11	0.84	1.20	1.34	0.22	1.46	1.24
Zenga	**4.19**	**2.88**	**2.55**	3.90	2.61	0.89	2.08	1.80	0.35	1.34	6.33
$GE_{1/2}$	2.32	2.22	1.69	2.17	1.52	0.55	0.98	1.02	0.50	0.49	4.13
T	1.71	1.59	1.21	0.99	0.68	0.31	0.31	0.38	0.10	0.34	4.74
GE_2	0.47	0.39	0.34	0.14	0.12	0.07	0.01	0.03	0.01	0.06	3.77
$R_{1/2}$	2.28	2.16	1.67	2.14	1.50	0.54	0.96	1.01	0.50	0.48	3.00
R_2	1.18	1.03	0.73	0.32	0.23	0.14	0.05	0.08	0.01	0.14	8.45
IQR	0.00	0.01	0.38	3.43	2.27	0.95	0.00	1.39	0.01	0.01	**0.00**
IDR	0.08	0.01	0.71	3.72	2.59	0.84	0.84	2.25	0.00	0.01	0.01
MQR	0.12	0.06	0.45	**4.70**	3.06	**1.18**	0.77	2.03	0.14	**5.97**	**0.00**

Principaux résultats I

- Robustesse :
  - Performance vs robustesse.
  - Importance des mesures basées sur les quantiles et les *quantile shares*.
- Trois types de mesures détectent mal les changements :
  - Indices basés sur l'entropie.
  - Rapports simples de quantiles.
  - Coefficient de variation.
- Bonne performance des indices basés sur les *quantile shares*.

Principaux résultats II

- Les indices les plus performants :
  - **Zenga :** très polyvalent, mais très sensible aux valeurs extrêmes.
  - **Atkinson ($\alpha = 1$) :** peu robuste et problème des revenus nuls.
  - **ASR :** meilleure mesure basée sur les quantile shares (robuste).
- Globalement, les indicateurs de Laeken (Gini et *QSR*) donnent des résultats plutôt acceptables, surtout *QSR*.
- Pour des mesures basées sur les queues de distribution, *QSR* et *DSR* se montrent performants et robustes.
- *MQR* est dans l'ensemble la meilleure mesure construite sur les rapports simples de quantiles.

Conclusion

- Compromis performance / robustesse.
- Importance des mesures utilisant les *quantile shares*.
- Importance des mesures nouvelles (*ASR*, *MQR*) ou récentes (Zenga).
- Perspectives : estimation dans des plans de sondage complexes.