ANNEXE 1-F : DÉVELOPPEMENT DE LA FORMULE DE CALCUL DE L'INDICE DE GINI

Dans le cas où le nombre d'observations est fini (cas discret), la différence moyenne de Gini s'écrit [1] :

$$\Delta = \frac{1}{N^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \left| y_j - y_k \right| f_j \, f_k$$

où

n est le nombre de valeurs distinctes observées

où f_j est la fréquence de la valeur y_j dans la distribution, de sorte que

$N = \sum_{j=1}^{n} f_j$ est le nombre d'observations

Lorsque les observations sont groupées par classes, la valeur y_j est la valeur moyenne de la variable Y dans la classe j.

Écrivons

$v_j = \frac{f_j}{N}$, la fraction de la population appartenant à la classe j.

La valeur moyenne de la variable Y s'écrit alors

$$\mu = \frac{1}{N} \sum_{j=1}^{n} f_j \, y_j = \sum_{j=1}^{n} v_j y_j$$

Soit

$M = \sum_{j=1}^{n} f_j y_j$, la somme des valeurs de la variable Y, et

$w_j = \frac{f_j y_j}{\sum_{k=1}^{n} f_k y_k} = \frac{f_j y_j}{N\mu} = \frac{v_j y_j}{\mu}$, la fraction de la somme allouée à la classe j.

---

[1] Dans cette formule chaque observation est comparée à chacune des observations, y compris à elle-même ; c'est la différence moyenne *avec* répétition. Kendall et Stuart (1991, p. 58) donnent aussi la formule *sans* répétition. Lorsque N est grand, la différence est négligeable.

Développons maintenant la formule de la différence moyenne de Gini :

$$\Delta = \frac{1}{N^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \left| y_j - y_k \right| f_j\, f_k$$

$$\Delta = \sum_{j=1}^{n} \sum_{k=1}^{n} \left| y_j - y_k \right| v_j v_k$$

$$\Delta = \sum_{j=1}^{n} \sum_{k=1}^{n} \left| v_j v_k y_j - v_j v_k y_k \right|$$

$$\Delta = \sum_{j=1}^{n} \sum_{k=1}^{n} \left| v_k\, \mu w_j - v_j\, \mu w_k \right|$$

$$\Delta = \mu \sum_{j=1}^{n} \sum_{k=1}^{n} \left| v_k\, w_j - v_j\, w_k \right|$$

La suite du développement s'appuie sur deux observations. D'abord,

$v_k\, w_j - v_j\, w_k = 0$ pour $k = j$.

Ensuite, si les observations ont été rangées, en vue de la construction d'une courbe de Lorenz, par ordre croissant des rapports w_j / v_j , alors

$$[k < j] \Rightarrow \left[{}^{w_j}\!/_{v_j} \geq {}^{w_k}\!/_{v_k} \right] \Rightarrow \left[v_k w_j - v_j w_k \geq 0 \right]$$

et on peut écrire

$$\Delta = 2\mu \sum_{j=1}^{n} \sum_{k<j} \left| v_k\, w_j - v_j\, w_k \right| = 2\mu \sum_{j=1}^{n} \sum_{k<j} \left(v_k\, w_j - v_j\, w_k \right)$$

$$\Delta = 2\mu \left(\sum_{j=1}^{n} \sum_{k<j} v_k\, w_j - \sum_{j=1}^{n} \sum_{k<j} v_j\, w_k \right)$$

$$\Delta = 2\mu \left(\sum_{j=1}^{n} w_j \sum_{k<j} v_k - \sum_{j=1}^{n} v_j \sum_{k<j} w_k \right)$$

Développons le premier des deux termes entre parenthèses :

$$\sum_{j=1}^{n} w_j \sum_{k<j} v_k = w_1 v_0 + w_2 (v_1) + w_3 (v_1 + v_2) + w_4 (v_1 + v_2 + v_3) + \ldots$$

où $v_0 = 0$.

$$\sum_{j=1}^{n} w_j \sum_{k<j} v_k = \sum_{k=1}^{n} v_k \sum_{j>k} w_j = \sum_{k=1}^{n} v_k \left(1 - \sum_{j \le k} w_j\right) = \sum_{k=1}^{n} v_k \left(1 - \sum_{j=1}^{k} w_j\right)$$

Pour finir, complétons la notation en posant

$$Cw_j = \sum_{k=1}^{j} w_j$$

On obtient alors

$$\sum_{j=1}^{n} w_j \sum_{k<j} v_k = \sum_{k=1}^{n} v_k \left(1 - Cw_k\right) = 1 - \sum_{k=1}^{n} v_k \, Cw_k = 1 - \sum_{j=1}^{n} v_j \, Cw_j$$

Et il résulte

$$\Delta = 2\mu \left(1 - \sum_{j=1}^{n} v_j \, Cw_j - \sum_{j=1}^{n} v_j \, Cw_{j-1}\right)$$

L'indice de concentration de Gini est simplement le rapport de la différence moyenne de Gini sur deux fois la moyenne :

$$G = \frac{\Delta}{2\mu} = 1 - \left(\sum_{j=1}^{n} v_j \, Cw_j + \sum_{j=1}^{n} v_j \, Cw_{j-1}\right) = 1 - \sum_{j=1}^{n} v_j \left(Cw_j + Cw_{j-1}\right)$$