원의 성질을 이용한 Lorenz 곡선과 Gini index의 추정

한 준 태[1], 강 석 복[2], 조 영 석[3]

요약문

소득분배의 가장 대표적인 불평등척도는 Gini index이며, 이것은 통계학자인 Gini가 제안한 지표로서 소득분배에 관한 분석에서 가장 널리 이용되고 있다. 본 논문에서는 두 원의 호에 의해 Lorenz 곡선을 추정하고 코사인법칙을 이용하여 Gini index를 추정하기 위한 새로운 간편한 방법을 제시하여, 소득분포를 따르는 파레토분포에서 모의실험을 통해 Ogwang and Rao (1996)의 추정방법과 평균제곱오차 면에서 비교 분석한다.

주요용어: Gini index, Lorenz 곡선, 원

1. 서론

Lorenz 곡선(LC)은 경제학분야에서 소득분배의 불균형 정도에 대한 척도로 널리 이용되는 곡선으로 사람들을 소득 순으로 나열할 때 총 인구에 대한 주어진 수준보다 소득이 적은 인구의 누적비율을 x축에 나타내고 이 사람들에 의한 총소득에 대한 소득의 누적비율을 y축에 나타낼 때 다양한 수준에서 인구의 누적비율과 소득의 누적비율의 값들을 연결하여 얻을 수 있는 곡선이다. 이 Lorenz 곡선을 수학적으로 표시하면

$$L(y)=\int_0^y x\ dF(x)/E(Y)$$

이고, 여기서 Y는 기대값 $E(Y)$가 존재하는 음이 아닌 소득변수이며, $F(y)$는 Y의 누적분포함수(cdf)이다. 일반적으로 $F^{-1}(p)=\inf_x\{x:F(x)\geq p\}$로 정의할 때 Gastwirth (1971) 는 Lorenz 곡선을 다음과 같이 정의했다.

$$L(p)=\int_0^p F^{-1}(x)dx/E(Y)$$

Lorenz 곡선을 이용하여 소득분배 정도를 비교하는 것이 대단히 편리하다. 그러나 두 Lorenz 곡선들이 서로 교차하는 경우에는 직관적인 관찰만으로 어느 쪽이 더 소득분배가 균형을 이루는지 판단하기 어렵다. 이 때 Gini index를 구하여 비교할 수 있는데 Gini index 는 소득분배의 불균형에 관한 척도 중 시각적으로 가장 널리 이용되는 Lorenz 곡선과 대각선 사이 면적의 두 배로 정의한다. 1912년 처음 공식화한 이탈리아의 통계학자인 Gini 이후에 Gini index, Gini concentration ratio, Gini coefficient 등으로 불리고 있다. 높은 불균형 소득분배를 갖는 국가들에 대한 Gini index는 대체로 0.5와 0.7 사이인 반면 비교적 균등하게 분배되는 국가에 대해서는 0.2에서 0.3 사이이다. Gini index를 수학적으로 표시하면 다음과 같이 정의할 수 있다.

---

1 영남대학교 수학통계학부, (712-749) 경북 경산시 대동 214-1

2 영남대학교 수학통계학부 교수, (712-749) 경북 경산시 대동 214-1

3 밀양대학교 자율전공학부 전임강사, (627-702) 경상남도 밀양시 내이동 1025-1

$$G = 1 - 2\int_0^1 L(p)\, dp$$

Lorenz 곡선과 Gini index에 관한 연구는 Moothathu (1985, 1990)가 지수분포에서 Lorenz 곡선과 Gini index의 최대가능도추정량을 제시하고 그 분포를 연구하였으며, 1990년에는 파레토분포에서 Lorenz 곡선과 Gini index의 균일최소분산비편향추정량을 제시하였다. Kang and Cho (1997)는 소득분포로 널리 알려진 파레토분포의 Gini index의 여러 추정량을 제안하였으며, Kang and Cho (1999)는 파레토분포에서 Lorenz 곡선의 여러 통계량을 제시하고 이들을 평균제곱오차 면에서 비교하였다.

본 논문에서는 주어진 데이터에 근거하여 Lorenz 곡선을 적절한 원들의 호로 연결된 것으로 생각하여 삼각함수를 이용해 Gini index를 추정하는 방법을 소개하고 소득분포로 널리 알려진 파레토분포에서 모의실험을 통해 평균제곱오차 면에서 비교하여 제안된 추정방법이 효율적임을 보인다.

2. Lorenz 곡선과 Gini index의 추정

소득을 크기 순서로 나열할 때, 수평축에는 주어진 소득을 가진 사람들의 누적비율 p를 나타내고 수직 축에서는 총소득에 대한 그들의 소득 누적비율 y를 나타내는 점들 (p, y)에 의해 정의되는 Lorenz 곡선 $L(p) = y$는 다음 성질을 만족한다.

(a) $L(0) = 0$

(b) $L(1) = 1$

(c) $L(p) \le p$

(d) $L(p)$는 볼록함수이다.

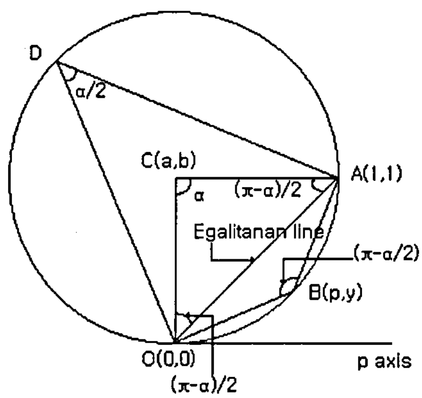

그림 1. 최적 원의 호로서 Lorenz 곡선

이러한 성질을 만족하는 함수로 그림 1과 같이 중심이 $C(a, b)$인 원의 호로 Lorenz 곡선을 추정하는 방법을 Ogwang and Rao (1996)가 제안하였다. 여기서 Lorenz 곡선과 Gini index의 추정은 다음과 같다.

그림 1과 같이 원의 호 OA 위의 한 점 $B(p, y)$를 생각하여 삼각형 OBA에서 제2코사인법칙을 이용하면

$$\cos B = \frac{OB^2 + BA^2 - OA^2}{2 \times OB \times BA}$$

이 성립된다. 또한 $OB^2 = p^2 + y^2$, $BA^2 = (p-1)^2 + (y-1)^2$, 그리고 $OA^2 = 2.0$이다.

주어진 관측값을 $B_1(p_1, y_1)$, $B_2(p_2, y_2), \ldots, B_n(p_n, y_n)$이라 하고 $\angle OB_iA \equiv B_i$를 코사인법칙을 이용하여 계산한 후 이 들의 표본평균 $\overline{B} = \sum_{i=1}^{n} B_i / n$을 $B(p, y)$에서의 $\angle OBA$의 추정값으로 사용한다면, $\angle OCA = \alpha = (2\pi - 2\overline{B})$이고 $\angle COA = \angle CAO = (\pi - \alpha)/2$이다. 삼각형 OAC에서 사인법칙을 이용하면

$$\frac{\sin \alpha}{OA} = \frac{\sin\left(\frac{\pi - \alpha}{2}\right)}{OC}$$

이므로 원의 반지름 $r(=OC)$은 다음과 같이 계산할 수 있다.

$$\begin{aligned} r &= \frac{\sqrt{2}\sin(\overline{B} - \pi/2)}{\sin(2\pi - 2\overline{B})} \\ &= \frac{1}{\sqrt{2}\sin(\pi - \overline{B})} \\ &= \frac{1}{\sqrt{2}\sin \overline{B}}. \end{aligned}$$

반지름이 r인 원에서 중심각이 α인 부채꼴의 넓이는 $r^2\alpha/2$이고, 삼각형 OAC의 넓이는 $(r^2 \sin \alpha)/2$이므로 Ogwang and Rao (1996)는 Gini index를 다음과 같이 계산했다.

$$\begin{aligned} G_O &= r^2(\alpha - \sin \alpha) \\ &= r^2[(2\pi - 2\overline{B}) - \sin(2\pi - 2\overline{B})]. \end{aligned}$$

일반적으로 소득분배에 불균형이 심한 경우는 Ogwang and Rao (1996)의 방법으로 추정할 경우 상당한 오차를 포함할 가능성이 크기 때문에 보완이 필요하다고 생각된다.

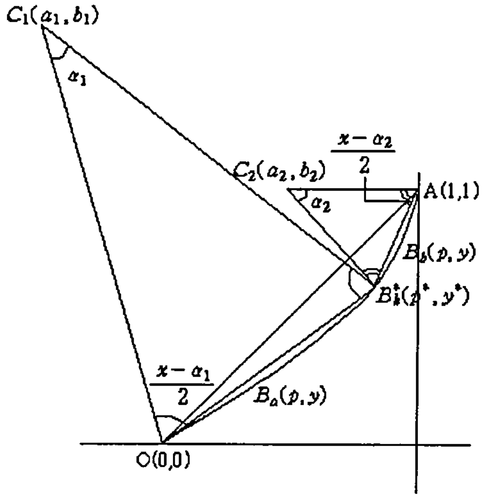

그림 2. 중심이 각각 $C_1(a_1, b_1)$과 $C_2(a_2, b_2)$인 두 원의 호로 생각한 Lorenz 곡선

이러한 경우는 Lorenz 곡선을 한 원의 호로 생각하기보다는 몇 개의 원들의 호로 이루어졌다고 생각하는 것이 훨씬 타당할 것이다. 관측값들 $B_1(p_1, y_1)$, ..., $B_2(p_2, y_2)$, $B_n(p_n, y_n)$에 대해 $\angle OB_iA$들이 비슷한 경우에는 Lorenz 곡선을 한 원의 호로 간주하여 Gini index를 추정해도 문제는 없으나 Lorenz 곡선이 치우침이 심한 경우는 이 곡선을 한 원의 호로 생각하기에는 적절하지 않기 때문에 그림 2과 같이 두 원을 이용하는 새로운 방법을 생각하였다.

각각의 관측값에 대한 $(p_i - y_i)$을 계산하여 최대인 점을 $B_k^*(p^*, y^*)$라 하면, 이 점을 기준으로 그림 3에서 호 OB_k^* 위의 점 $B_a(p, y)$와 호 B_k^*A 위의 점 $B_b(p, y)$를 생각하여 $\angle OB_aB_k^*$와 $\angle B_k^*B_bA$를 추정하고 이 각의 추정값을 이용해 반경을 계산한 다음 코사인법칙에 의해 Gini index를 추정한다. $\angle B_i$를 구하기 위해 제2코사인법칙을 이용하면

$$\cos B_i = \frac{OB_i^2 + B_iB^{*2} - OB^{*2}}{2 \times OB_i \times B_iB^*} \quad , i = 1, 2, \ldots, k-1 \tag{1}$$

$$\cos B_j = \frac{B^*B_j^2 + B_jA^2 - B^*A^2}{2 \times B^*B_j \times B_jA} \quad , j = k+1, \ldots, n \tag{2}$$

이다.

한 원의 현에서 원주각은 항상 동일하지만 사실 관측값 $B_1(p_1, y_1)$, ..., $B_{k-1}(p_{k-1}, y_{k-1})$이 모두 한 원의 호 위에 놓이지 않기 때문에 식 (1)을 이용하여 $\angle OB_iA,\ i = 1, \ldots, k-1$를 구한 다음 그들의 평균 $\overline{B_a} = \sum_{i=1}^{k-1} B_i/(k-1)$을 계산하여 $B_a(p, y)$에서 $\angle B_a$의 추정값으로 사용하고, $B_{k+1}(p_{k+1}, y_{k+1})$, ..., $B_n(p_n, y_n)$도 역시 또 다른 한 원의 호 위에 모두 놓이지 않기 때문에 식(2)를 이용하여 $\angle OB_jA,\ j = k+1, \ldots, n$을 구한 다음 그들의 평균 $\overline{B_b} = \sum_{j=k+1}^{n} B_j/(n-k)$를 계산하여 $B_b(p, y)$에서 $\angle B_b$의 추정값으로 사용한다. 그러면

$$\angle OC_1B_k^* = \alpha_1 = (2\pi - 2\overline{B_a}), \quad \angle C_1OB_k^* = \angle C_1B_k^*O = (\pi - \alpha_1)/2$$

이며,

$$\angle B_k^*C_2A = \alpha_2 = (2\pi - 2\overline{B_b}), \quad \angle C_2B_k^*A = \angle C_2AB_k^* = (\pi - \alpha_2)/2$$

이다. 삼각형 $OB_k^*C_1$에서 사인법칙을 이용하면

$$\frac{\sin\alpha_1}{OB_k^*} = \frac{\sin\left(\frac{\pi - \alpha_1}{2}\right)}{OC_1}$$

이므로 중심이 $C_1(a_1, b_1)$인 원의 반지름 $r_1(=OC_1)$은 다음과 같이 계산할 수 있다.

$$\begin{aligned} r_1 &= \frac{\sin(\pi/2 - \alpha_1/2)}{\sin\alpha_1} \times \overline{OB_k^*} \\ &= \frac{\sqrt{p^{*2} + y^{*2}}}{2\sin(\alpha_1/2)} \end{aligned}$$

$$= \frac{\sqrt{p^{*2} + y^{*2}}}{2\sin(\overline{B_a})}.$$

삼각형 $B_k^* A C_2$에서 사인법칙을 이용하면

$$\frac{\sin\alpha_2}{B_k^* A} = \frac{\sin\left(\frac{\pi - \alpha_2}{2}\right)}{B_k^* C_2}$$

이므로 중심이 $C_2(a_2, b_2)$인 원의 반지름 $r_2 (= B_k^* C_2)$는 다음과 같이 계산할 수 있다.

$$\begin{aligned} r_2 &= \frac{\sin(\pi/2 - \alpha_2/2)}{\sin\alpha_2} \times B_k^* A \\ &= \frac{\cos(\alpha_2/2) \times B_k^* A}{2\sin(\alpha_2/2)\cos(\alpha_2/2)} \\ &= \frac{\sqrt{(1-p^*)^2 + y(1-y^*)^2}}{2\sin(\overline{B_b})}. \end{aligned}$$

반지름의 길이가 r_1이고 중심각이 α_1인 부채꼴의 넓이는 $S_1 = \frac{1}{2} r_1^2 \alpha_1$이고 삼각형 $OB_k^* C_1$의 넓이는 $\frac{1}{2} r_1^2 \sin\alpha_1$이며, 반지름의 길이가 r_2이고 중심각이 α_2인 부채꼴의 넓이는 $S_2 = \frac{1}{2} r_2^2 \alpha_2$이고 삼각형 $B_k^* A C_2$의 넓이는 $\frac{1}{2} r_2^2 \sin\alpha_2$이다. 또한 삼각형 $OB_k^* A$의 넓이는 헤론의 법칙에 의해

$$\sqrt{S(S-\sqrt{2})(S-B_k^* A)(S-OB_k^*)}$$

이다. 여기서 선분

$$B_k^* A = \sqrt{(1-p^*)^2 + (1-y^*)^2},$$
$$OB_k^* = \sqrt{p^{*2} + y^{*2}}$$

이고

$$S = \frac{1}{2}(\sqrt{2} + B_k^* A + OB_k^*)$$

이다. 따라서 이 결과를 이용하여 새로운 Gini index의 추정량을 다음과 같이 제시할 수 있다.

$$\begin{aligned} G_K &= r_1^2(\alpha_1 - \sin\alpha_1) + r_2^2(\alpha_2 - \sin\alpha_2) \\ &\quad + 2\sqrt{S(S-\sqrt{2})(S-B^* A)(S-OB^*)}. \end{aligned}$$

참고문헌

[1] Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, Vol. 39, 1037-1038.

[2] Kang, S. B. and Cho, Y. S. (1997). Estimation of Gini index of the Pareto distribution. *Journal of Information & Optimization Sciences*, Vol. 18(3), 383-392.

[3] Kang, S. B. and Cho, Y. S. (1999). Estimation of the Lorenz curve of the Pareto distribution. *The Korean Communications in Statistics*, Vol. 6(1), 285-292.

[4] Moothathu, T. S. K. (1985). Distributions of Maximum Likelihood Estimators of Lorenz Curve and Gini Index of Exponential Distribution. *Ann. Inst. Statist. Math.*, Vol. 37, Part A, 473-479.

[5] Moothathu, T. S. K. (1990). The Best Estimator of Lorenz Curve, Gini Index and Theil Entropy Index of Pareto Distribution. *Sankhyā*, Vol. 52(B), 115-127.

[6] Ogwang T. and Rao, U. L. G. (1996). A new functional form for approximation the Lorenz curve, *Economics Letters*, Vol. 52, 21-29.