Vol. 9 - pp. 35-42
Junio - Noviembre 2005 - ISSN: 0329-5583
http://vaca.agro.uncor.edu/~revistasae/

NORMALIDAD ASINTÓTICA PARA LOS E-GINI

Pablo Martínez Camblor

UAMAC, Universidad Autónoma de Tamaulipas, Cd. Victoria, Tamaulipas, México

RESUMEN

En este trabajo se presenta una demostración de la normalidad asintótica para los E-Gini, familia que generaliza el popular índice de desigualdad propuesta por Chakravarty (1988).

El método de demostración utilizado se basa en el conocido Proceso Húngaro y es fácilmente generalizable a otros estadísticos que involucran funcionales de la Función de Distribución Empírica.

El resultado teórico obtenido se adapta para ser utilizado en la práctica mediante un método plug-in en el que se sustituyen las expresiones no calculables por estimaciones suaves basadas en la Función de Distribución Empírica Suavizada.

PALABRAS CLAVE: E-Gini, Proceso Húngaro, Estimación Núcleo.

1. INTRODUCCIÓN

El Índice de Concentración de Gini (ICG) es probablemente la medida de desigualdad de rentas más utilizada. Dada la naturaleza de las variables en las que se usa, es razonable suponer que su función de distribución, F, verifica que existen números reales $0 \leq m < M < \infty$ tales que, F $(m) = 0$ y F $(M) = 1$, quedando este definido por,

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

donde L es la función de Lorenz.

teniendo en cuenta que $0 < \mu = \int x dF(x) < \infty$, una expresión equivalente será

$$ICG = \frac{\int_m^M F(y)(1 - F(y))dy}{\mu} \qquad (2)$$

y, sin más que sustituir la función de distribución real por su función de distribución empírica, se tiene el estimador usado normalmente para el ICG,

$$\hat{G}_n = \frac{\int_m^M F_n(y)\left(1-F_n(y)\right)dy}{\overline{X}} \tag{3}$$

Existen distintas generalizaciones para el ICG. Donalson y Weymark (1980, 1983) y Yitzhaki (1983) definen los S-Gini para cada $\eta > 1$ como,

$$SG_n = \frac{1}{\mu}\left(M-(k-1)m-\int_m^M \left(F(y)+(k-1)\left(1-F(y)\right)^{\eta}\right)dy\right)$$

Notar que $SG2 = ICG$ y, sustituyendo la función de distribución real por la función de distribución empírica, nuevamente se obtiene el estimador usualmente utilizado,

$$\widehat{SG}_{\eta} = \frac{1}{\overline{X}}\left(M-(\eta-1)m-\int_m^M \left(F_n(y)+(n-1)\left(1-F_n(y)\right)^{\eta}\right)dy\right)$$

Otra generalización del Gini son los E-Gini (Chakravarty, 1988) definidos para cada $\delta \geq 1$ como

$$EG_{\delta} = \frac{2}{\mu}\left(\int_m^M \left(\mu F(y)-y\left(1-F(y)\right)^{\delta}\right)dy\right)^{1/\delta}$$

nuevamente, notar que $EG1 = ICG$ y que su estimador natural viene dado por,

$$\widehat{EG}_{\delta} = \frac{2}{\overline{X}}\left(\int_m^M \left(\overline{X}F_n(y)-y\left(1-F_n(y)\right)^{\delta}\right)dy\right)^{1/\delta}$$

En este trabajo, partiendo de una muestra aleatoria simple $x_1, \ldots, x_n$ se demuestra la distribución asintótica del estimador $\widehat{EG}_{\delta}$.

Como suele ser usual en este tipo de problemas, tanto la media como la varianza dependen de la distribución real de la variable en estudio por lo que, para su uso en la práctica se reemplazará la función de distribución real por su Función de Distribución Empírica Suavizada (Nadaraya, 1964) y que, dado una muestra aleatoria $x_1, \ldots, x_n$ y $t \in \mathbb{R}$ queda definida por,

$$\widetilde{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{K}\left(\frac{x_i - t}{h}\right) \tag{8}$$

donde $\widetilde{K}(u) = \int_{-\infty}^{u} K(p)\,dp$, y K es una función de densidad usualment simétrica respecto al origen y $h = h_n$ una sucesión de números positivos.

2. APROXIMACIONES ASINTÓTICAS

En este apartado se demostrará la normalidad asintótica para el estadístico (7). El método de demostración usado, se basa en el conocido como proceso húngaro, dado por Kómlos, Mayor y Tusnády (1975) y que, garantiza la existencia de un espacio probabilístico y una sucesión de puentes brownianos, $W^0 = W_n^0$, de modo que,

$$\sqrt{n}\left(F_n(X,t)-F(t)\right)=W^0\{F(t)\}+\frac{\log n}{\sqrt{n}}Z_n(t)$$

donde $\sup_{t\in\mathbb{R}}|Z_n(t)|$ es una variable aleatoria casi seguro acotada.

y consiste en dejar la expresión cuya normalidad asintótica se quiere demostrar como un funcional de $\sqrt{n}\left(F_n(X,t)-F(t)\right)$ aplicar la ecuación y las p de los puentos brownianos.

Para mayor simplicidad en las expresiones, $\int$ represen $\int_m^M$.

Teorema: Dada una muestra aleatoria simple $x_1, \ldots, x_n$, procedente de una variable aleatoria absolutamente continua cuya función de densidad tiene soporte compacto, entonces,

$$\sqrt{n}\,\frac{\widehat{EG_\delta}-EG_\delta}{\nu_\delta}\xrightarrow{\mathfrak{L}}\mathcal{N}(0,1) \tag{10}$$

donde,

$$\mathcal{V}_\delta^2=\iint\mathfrak{F}(y)\mathfrak{F}(x)\left(\min\{F(x),F(y)\}-F(x)F(y)\right)dxdy$$

siendo

$$\mathfrak{F}(u)=\frac{2^\delta}{(EG_\delta)^{\delta-1}\mu^\delta}\left((\mu+y)\left(\mu F(y)-y(1-F(y))^{\delta-1}-\left(\frac{EG_\delta}{2}\right)^\delta\right)\right)$$

Demostración

En primer lugar, $\bar{X}=\mu+O_p\left(n^{-1/2}\right), F_n(u)=F(u)+O_p\left(n^{-1/2}\right)$

más que tener en cuenta que se da la igualdad,

$$\sqrt{n}\left(\left(\widehat{EG_\delta}\right)^{\delta}-\left(EG_\delta\right)^{\delta}\right)=\left(EG_\delta\right)^{\delta-1}\sqrt{n}\left(\left(\frac{\widehat{EG_\delta}}{EG_\delta}\right)^{\delta-1}\widehat{EG_\delta}-EG_\delta\right)=$$

$$\left(EG_\delta\right)^{\delta-1}\sqrt{n}\left(\left(\frac{EG_\delta+O_p\left(n^{-1/2}\right)}{EG_\delta}\right)^{\delta-1}\widehat{EG_\delta}-EG_\delta\right)=$$

$$\left(EG_\delta\right)^{\delta-1}\sqrt{n}\left(\widehat{EG_\delta}-EG_\delta\right)+O_p\left(n^{-1/2}\right)$$

desde la igualdad anterior y las definiciones (6) y (7),

$$\sqrt{n}\left(\widehat{EG_\delta}-EG_\delta\right)=\sqrt{n}\frac{2^{\delta}}{\left(EG_\delta\right)^{\delta-1}\mu^{\delta}}\left(\int\left(\bar{X}F_n(y)-y\left(1-F_n(y)\right)^{\delta}\right)dy\right)$$

$$-\sqrt{n}\frac{2^{\delta}}{\left(EG_\delta\right)^{\delta-1}\mu^{\delta}}\left(\left(\frac{\bar{X}}{\mu}\right)^{\delta}\int\left(\mu F(y)-y\left(1-F(y)\right)^{\delta}\right)dy\right)+O_p\left(n^{-1/2}\right)$$

Obviamente,

$$\sqrt{n}\left(\int\left(\bar{X}F_n(y)-y\left(1-F_n(y)\right)^{\delta}\right)dy-\left(\frac{\bar{X}}{\mu}\right)^{\delta}\int\left(\mu F(y)-y\left(1-F(y)\right)^{\delta}\right)dy\right)=$$

$$\sqrt{n}\left(\int\left(\bar{X}F_n(y)-y\left(1-F_n(y)\right)^{\delta}\right)dy-\int\left(\mu F(y)-y\left(1-F(y)\right)^{\delta}\right)dy\right)$$

$$+\sqrt{n}\left(1-\left(\frac{\bar{X}}{\mu}\right)^{\delta}\right)\int\left(\mu F(y)-y\left(1-F(y)\right)^{\delta}\right)dy$$

(14)

Teniendo en cuenta nuevamente que $F_n(u)=F(u)+O_p\left(n^{-1/2}\right)$ $\bar{X}=\mu+O_p\left(n^{-1/2}\right)$ se tiene que,

$$\sqrt{n}\left(\bar{X}F_n(y) - y\left(1-F_n(y)\right)^{\delta}\right) =$$
$$\left(\bar{X}F_n(y) - y\left(1-F_n(y)\right)^{\delta-1}\sqrt{n}\left(\bar{X}F_n(y) - y\left(1-F_n(y)\right)\right)\right) =$$
$$\left(\bar{X}\left(F(y)+O_p\left(n^{-1/2}\right)\right)\right)$$
$$-y\left(1-\left(F(y)+O_p\left(n^{-1/2}\right)\right)^{\delta-1}\right)\sqrt{n}\left(\bar{X}F_n(y)\right)$$
$$-y\left(1-F_n(y)\right) =$$
$$\left(\left(\mu+O_p\left(n^{-1/2}\right)\right)F(y) - y\left(1-F(y)\right)^{\delta-1}\sqrt{n}\left(\bar{X}F(y) - y\left(1-F_n(y)\right)\right)\right)$$
$$+O_p\left(n^{-1/2}\right) =$$
$$\left(\mu F(y) - y\left(1-F(y)\right)^{\delta-1}\left(\left(\mu+O_p\left(n^{-1/2}\right)\right)\right)\right)\sqrt{n}F_n(y) - y\sqrt{n}\left(1-F(y)\right)$$
$$+O_p\left(n^{-1/2}\right) =$$
$$\left(\mu F(y) - y\left(1-F(y)\right)^{\delta-1}\right)\left(\mu\sqrt{n}F_n(y) - y\sqrt{n}\left(1-F_n(y)\right)+O_p\left(n^{-1/2}\right)\right)$$

Por otro lado,

$$\sqrt{n}\left(\mu F(y)\right) - y\left(1-F(y)\right)^{\delta} =$$
$$\left(\mu F(y) - y\left(1-F(y)\right)^{\delta-1}\left(\mu\sqrt{n}F(y) - y\sqrt{n}\left(1-F(y)\right)\right)\right)$$

luego,

$$\sqrt{n}\left(\int\left(\bar{X}F_n(y) - y\left(1-F_n(y)\right)^{\delta}\right)dy\right) - \int\left(\mu F(y) - y\left(1-F(y)\right)^{\delta}\right)dy =$$
$$\int\left(\mu F(y) - y\left(1-F(y)\right)^{\delta-1}\left(\mu\sqrt{n}\left(F_n(y)\right)\right)\right)dy$$
$$-\int\left(\mu F(y) - y\left(1-F(y)\right)^{\delta-1} y\sqrt{n}\left(\left(1-F_n(y)\right)-\left(1-F(y)\right)\right)\right)dy =$$
$$\int(\mu+y)\left(\mu F(y) - y\left(1-F(y)\right)^{\delta-1}\sqrt{n}\left(F_n(y)-F(y)\right)\right)dy$$

Sin más que aplicar el teorema de integración por partes en el segundo sumando de (14) se obiene la igualdad,

$$\sqrt{n}\left(1-\left(\frac{\overline{X}}{\mu}\right)^{\delta}\right)\int\left(\mu F(y)-y\left(1-F(y)\right)^{\delta}\right)dy=$$

$$\frac{1}{\mu^{\delta}}\int\left(\mu F(y)-y\left(1-F(y)\right)^{\delta}\right)dy\sqrt{n}\int xd\left[F(x)-F_{n}(x)\right]=$$

$$\left(\frac{EG_{\delta}}{2}\right)^{\delta}\int\sqrt{n}\left(F_{n}(y)-F_{n}(y)\right)dy$$

Desde (13), (14), (17), (18), y la definición de $\mathcal{F}$ se tiene,

$$\sqrt{n}\left(\widehat{EG_{\delta}}-EG_{\delta}\right)=\int\mathcal{F}(y)\sqrt{n}\left(F_{n}(y)-F(y)\right)dy+O_{p}\left(n^{1/2}\right)$$

Por otro lado, se tiene sin más que aplicar el teorema 3 de Kómlos, Mayor y Tusnády (1975) que existirá un espacio probabilístico y puentes brownianos, W^0 de modo que,

$$\int\mathcal{F}(y)\sqrt{n}\left(F_{n}(y)-F(y)\right)dy=\int\mathcal{F}(y)\mathcal{W}^{0}\left\{F(y)\right\}dy+O_{p}\left(n^{-1/2}\right)$$

luego si se define la variable aleatoria

$$Q_{n}^{\delta}=\int\mathcal{F}(y)\mathcal{W}^{0}\left\{F(y)\right\}dy \tag{21}$$

se tiene, sin más que aplicar el teorema del valor medio del cálculo integral, que existe un $m\leq\theta_{n}\leq M$ de modo que,

$$Q_{n}^{\delta}=\mathcal{F}(\theta_{n})\mathcal{W}^{0}\left\{F(\theta_{n})\right\} \tag{22}$$

de donde Q_{n}^{δ} tiene distribución normal de esperanza nula y varianza,

$$E\left[\left(Q_{n}^{\delta}\right)^{2}\right]=E\left(\int\mathcal{F}(y)\mathcal{W}^{0}\left\{F(y)\right\}dy^{2}\right)=$$

$$E\left(\int\int\mathcal{F}(y)\mathcal{F}(x)\mathcal{W}^{0}\left\{F(x)\right\}\mathcal{W}^{0}\left\{F(y)\right\}dydx\right)=$$

$$\int\int\mathcal{F}(y)\mathcal{F}(x)E\left(\mathcal{W}^{0}\left\{F(x)\right\}\mathcal{W}^{0}\left\{F(y)\right\}\right)dydx=$$

$$\int\int\mathcal{F}(y)\mathcal{F}(x)\left(\min\left\{F(x),F(y)\right\}-F(x)F(y)\right)dxdy=\mathcal{V}_{\delta}^{2}$$

Desde las igualdades,

$$\sqrt{n}\,\frac{\widehat{EG}_\delta - EG_\delta}{\mathcal{V}_\delta} = \frac{\mathcal{Q}_n^\delta + O_p\left(n^{-1/2}\right)}{\mathcal{V}_\delta} = \left(\mathcal{V}_\delta\right)^{-1}\mathcal{Q}_n^\delta + O_p\left(n^{-1/2}\right)$$

se deduce inmediatamente el resultado (10).

El funcional $\mathcal{F}$ depende de de la función de distribución real por ello, este resultado no puede ser utilizado en la construcción de intervalos de confianza para los EG_δ. Sin embargo, sin más que aplicar un método plug-in, sustituyendo la función de distribución real por su función de distribución empírica o, si se asumen ciertas condiciones de regularidad, por la función de distribución empírica suavizada dada en (8) se tendrá que si se definen,

$$\widehat{\mathcal{F}}(u) = \frac{2^\delta}{\left(\widehat{EG}_\delta\right)^{\delta-1}\bar{X}^\delta}\left(\left(\bar{X}+y\right)\left(\bar{X}F_n(y) - y\left(1-F_n(y)\right)^{\delta-1} - \left(\frac{\widehat{EG}_\delta}{2}\right)^\delta\right)\right)$$

Desde las leyes fuertes de los grandes números se tiene la convergencia,

$$\widehat{\mathcal{F}}(u) \xrightarrow{\mathcal{P}} \mathcal{F}(u)$$

luego si $\hat{\mathcal{V}}_\delta^2 = \int\int \widehat{\mathcal{F}}(y)\widehat{\mathcal{F}}(x)\left(\min\{F_n(x),F_n(y)\} - F_n(x)F(y)\right)dxdy$
la convergencia,

$$\sqrt{n}\,\frac{\widehat{EG}_\delta - EG_\delta}{\hat{\mathcal{V}}_\delta} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \tag{25}$$

Bajo condiciones de regularidad se tiene que,

$$\widetilde{\mathcal{F}}(u) \xrightarrow{\mathcal{P}} \mathcal{F}(u)$$

que definiendo,

$$\widetilde{\mathcal{V}}_\delta^2 = \int\int \widetilde{\mathcal{F}}(y)\widetilde{\mathcal{F}}(x)\left(\min\{\widetilde{F}_n(x),\widetilde{F}_n(y)\} - \widetilde{F}_n(x)\widetilde{F}_n(y)\right)dxdy$$

gencia,

$$\sqrt{n}\,\frac{\widehat{EG}_\delta - EG_\delta}{\widetilde{\mathcal{V}}_\delta} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \tag{26}$$

REFERENCIAS

Chakravarty, S. R. (1988) Extended Gini indices of inequality. *International Economic Review,* 29, 147–156.

Donaldson, D. y Weymark, J. A. (1980) A Single Parameter Generalization of the Gini Index And Inequality. *Journal of Economic Theory*, 22, 67–86.

Donaldson, D. y Weymark, J. A. (1983) Ethical Flexible Indices for Income Distributions in the Continuum. *Journal of Economic Theory,* 29, 353–358.

Giles, E. A. D. (2004) Calculating a Standard Error for the Gini Coefficient: Some Further Results. *Oxford Bulletin of Economics & Statistics,* Vol. 66, 425–428.

Gini, C. (1995) *Variabilit*á e *mutabilita*. 1912. Reprinted in Memorie di metodologia statistica (Ed. E. Pizetti and T. Salvemini.) Rome: Libreria Eredi Virgilio Veschi.

Glasser, G. J. (1962) Variance Formulas for the Mean Difference and Coefficient of Concentration. *Journal of American Statistics Assoc.*, 57, 648–654.

Komlós, J., Major, P. y Tusnády, G. (1975) An approximation of partial sums of independent RV-S, and the sample DF.I. Z. *Wahrscheinlichkeitstheor.* Verw. Geb., 32, 1-2, pp. 111–131.

Lerman, R. I. y Yitzhaki, S. (1984) A note on the calculation and interpretation of the Gini index. *Economics Letters,* Vol. 5, pp. 363-368.

Nadaraya, E. A. (1964) Some new estimates for distribution functions. *Theory Probab.* Applic. 9, 497-500.

Sen, A. (1973) *On Economic Inequality.* Oxford, England: Clarendon Press.

Yitzhaki, S. (1983) On an Extension of the Gini Inequality Index. *International Economic Review,* 24, 183–1.