

## Central Limit Theorems for S-Gini and Theil Inequality Coefficients

Teoremas central del límite para el S-Gini y el coeficiente de Theil

PABLO MARTÍNEZ-CAMBLOR<sup>a</sup>

FUNDACIÓN CAUBET-CIMERA ILLES BALEARS, MALLORCA, ESPAÑA

---

### Abstract

The Hungarian Construction (Kömlös et al. 1975) is used for getting a proof of asymptotic normality of S-Gini coefficient; this method is very interesting because it can be used to check asymptotic normality of other income inequality measures as Theil coefficient. Besides, explicit expressions of asymptotic means and variances are given for S-Gini and Theil estimators. Finally, to illustrate the performance of obtained results, we carry out a simulation study comparing the asymptotic and Smoothed Bootstrap approximations.

**Key words:** S-Gini index, Theil index, Hungarian construction, Kernel density estimation.

### Resumen

Se usa el Proceso Húngaro (Kömlös et al. 1975) para derivar la normalidad asintótica del S-Gini; Este método es muy interesante ya que puede ser usado para demostrar la normalidad asintótica de otros coeficientes usados para medir la desigualdad de ingresos como el de Theil. Se consiguen expresiones explícitas para la media y la varianza del S-Gini y del coeficiente de Theil. Finalmente, se realiza un estudio de simulación, en el que se compara el rendimiento de la aproximación asintótica propuesta y del método Bootstrap Suavizado.

**Palabras clave:** índice S-Gini, índice de Theil, proceso húngaro, estimación *kernel* para la densidad.

---

<sup>a</sup>Programa de epidemiología e investigación clínica. E-mail: martinez@caubet-cimera.es

## 1. Introduction

The Gini Concentration Index (Gini 1995) has been extensively used in the study of distribution inequality. If  $L$  is the Lorenz function it is defined as:

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

or, for certain random sample  $x_1, \dots, x_N$  and, if  $x_{(1)}, \dots, x_{(N)}$  are the sorted sample:

$$G = \sum_{i=1}^N \frac{(2i - N)x_{(i)}}{N \sum_{i=1}^N x_i}$$

In the general case, the variable in study is defined on the real interval  $(m, M)$  with  $0 \leq m < M < \infty$  (we always assume this condition), an equivalent expression for the Gini index used in Giles (2004) is

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

Replacing the distribution function for its maximum likelihood estimation (FDE),  $F_n$ , and if  $\bar{X} = \int x dF_n(x)$  we have the usual  $G$  estimator

$$\hat{G}_n = \frac{\int_m^M F_n(y)(1 - F_n(y)) dy}{\bar{X}}$$

This index has been very actively investigated for the last three decades. Its exact sample distribution in the particular case of a skew normal distribution has been studied by Crocetta & Loperfido (2005) under a more general case of the  $L$ -statistics. In the general case, its asymptotic distribution and the asymptotic distribution of other families which generalizes the Gini Index (E-Gini) have been studied by Zitikis (2003) and Martínez-Camblor (2005). The multivariate case has been studied by Martínez-Camblor (2007).

Donalson & Weymark (1980, 1983) and Yitzhaki (1983) propose the Single Parameter Gini (S-Gini) define as:

$$SG_k = 1 - k(k-1) \int (1-p)^{k-2} L(p) dp$$

$$= \frac{1}{\mu} \left( M - (k-1)m - \int (F(y) + (k-1)(1-F(y))^k) dy \right), \quad k > 1 \quad (1)$$

for  $k = 2$  we obtain the Gini standard coefficient.

In section 2 we derive the asymptotic normality of  $\hat{G}_n$  and using the same technique, we also get to prove the asymptotic distribution for S-Gini and Theil Coefficients (Theil 1967).

With a plug-in method, the main results are adapted to be used in practice. We replace both unknown expected value and variance for their respective smoothed

estimators, those are obtained when we replace the real distribution functions for the Smoothed Empirical Distribution Function (SEDF) defined by Nadaraya (1964) as:

$$\tilde{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \tilde{K} \left( \frac{t - x_i}{h_n} \right) \tag{2}$$

where  $\tilde{K}(t) = \int_{-\infty}^t K(s) ds$ ,  $K$  is a symmetric density function and  $h_n$  is a sequence of real positive number.

Section 3 is devoted to propose a resample method for  $\hat{G}_n$ . In this context the more useful technique is the Smoothed Bootstrap (Hall et al. 1989), we describe it in this concrete case.

Finally and in order to study the performance of two proposed methods, we show the obtained results in simulation work.

## 2. Main Results and Proofs

In this section we prove the main results of this paper. To derive them, we will be based on theorem 3 of Komlós et al. (1975) which imply that there exists a probabilist space  $(\Omega, \sigma, \mathcal{P})$  and a Brownian Bridge,  $W^0$ , such that

$$\sqrt{n}(F_n(x) - F(x)) = W^0\{F(x)\} + \frac{\log n}{\sqrt{n}} Z_n(x), \quad \text{a.s. (almost surely)}$$

where  $Z_n(x)$  is an random variable almost surely bounded.

A Brownian Bridge is a gaussian process with expected value,  $E(W^0\{t\}^2) = t(1-t)$  and  $E(W^0\{t\}W^0\{s\}) = s(1-t)$  for  $s \leq t$  (see, for example Billingsley 1968).

**Theorem 1.** *Let  $(x_1, \dots, x_n)$  a random sample from  $F$ , then*

$$\sqrt{n} \frac{\hat{G}_n - G}{\mathcal{V}} \xrightarrow{\mathcal{L}}_n \mathcal{N}(0, 1) \tag{3}$$

where

$$G = \frac{1}{\mu} \int F(u)(1 - F(u)) du \tag{4}$$

$$\begin{aligned} \mathcal{V}^2 = & \frac{1}{\mu^2} \int \int_m^u (1 - 2F(u) + G)(1 - 2F(v) + G)F(v)(1 - F(u)) dv du + \\ & \frac{1}{\mu^2} \int \int_u^M (1 - 2F(u) + G)(1 - 2F(v) + G)F(u)(1 - F(v)) dv du \end{aligned} \tag{5}$$

**Proof.** For each  $n \in \mathbb{N}$  we define

$$\xi_n = \sqrt{n} \left( \int F_n(x)(1 - F_n(x)) dx - \frac{\int x dF_n(x)}{\mu} \int F(x)(1 - F(x)) dx \right)$$

from the equality;  $F_n(x) = F(x) + O_{\mathcal{P}}(n^{-1/2})$ , and the Theorem 3 of Komlós et al. (1975) we have that

$$\begin{aligned} \xi_n &= \int \sqrt{n}(F_n(x) - F(x))(1 - F_n(x)) dx - \int F(x)\sqrt{n}(F_n(x) - F(x)) dx \\ &\quad + \sqrt{n} \int F(x)(1 - F(x)) dx \\ &\quad - \frac{1}{\mu} \int F(x)(1 - F(x)) dx \int x d\sqrt{n}(F_n(x) - F(x)) \\ &\quad - \frac{\sqrt{n}}{\mu} \int F(x)(1 - F(x)) dx \int x dF(x) \stackrel{\text{a.s.}}{=} \int \mathcal{W}^0\{F(x)\}(1 - F(x)) dx \\ &\quad + O_{\mathcal{P}}(n^{-1/2}) - \int \mathcal{W}^0\{F(x)\}F(x) dx \\ &\quad + O(\log n/\sqrt{n}) - \frac{1}{\mu} \int F(x)(1 - F(x)) dx \int x d\mathcal{W}^0\{F(x)\} + O(\log n/\sqrt{n}) \end{aligned}$$

applying the integration by parts,

$$\xi_n = \int (1 - 2F(x) + G)\mathcal{W}^0\{F(x)\} dx + O_{\mathcal{P}}(n^{-1/2}) = \xi + O_{\mathcal{P}}(n^{-1/2}) \quad \text{a.s.}$$

Now,  $F$  is vanish out of  $(m, M)$  so, we know that there exists  $\theta \in (m, M)$  such that

$$\xi = \int (1 - 2F(x) + G)\mathcal{W}^0\{F(x)\} dx = (M - m)(1 - 2F(\theta) + G)\mathcal{W}^0\{F(\theta)\} \quad (6)$$

has a normal distribution with mean zero and variance,

$$\begin{aligned} E(\xi^2) &= E \left[ \left( \int (1 - 2F(x) + G)\mathcal{W}^0\{F(x)\} dx \right)^2 \right] \\ &= E \left( \iint (1 - 2F(x) + G)(1 - 2F(y) + G)\mathcal{W}^0\{F(x)\}\mathcal{W}^0\{F(y)\} dx dy \right) \\ &= \iint (1 - 2F(x) + G)(1 - 2F(y) + G)E(\mathcal{W}^0\{F(x)\}\mathcal{W}^0\{F(y)\}) dx dy \end{aligned}$$

and from basic properties of Brownian Bridge we obtain the final expression for the variance

$$\begin{aligned} E(\xi^2) &= \iint_m^u (1 - 2F(u) + G)(1 - 2F(v) + G)F(v)(1 - F(u)) dv du + \\ &\quad \iint_u^M (1 - 2F(u) + G)(1 - 2F(v) + G)F(u)(1 - F(v)) dv du = (\mu\mathcal{V})^2 \end{aligned}$$

On the other hand,  $\bar{X} = \mu + O_{\mathcal{P}}(n^{-1/2})$ , so we have that

$$\sqrt{n} \frac{\widehat{G}_n - G}{\mathcal{V}} \stackrel{\text{a.s.}}{=} \frac{\xi_n}{\bar{X}\mathcal{V}} = \frac{\xi + O_{\mathcal{P}}(n^{-1/2})}{(\mu + O_{\mathcal{P}}(n^{-1/2}))\mathcal{V}} = \frac{\xi}{\mu\mathcal{V}} + O_{\mathcal{P}}(n^{-1/2}) \quad (7)$$

since  $\xi$  has a  $\mathcal{N}(0, \mu\mathcal{V})$  distribution, the proof is completed.  $\square$

This technique can be applied to prove the asymptotic normality of other similar rates. For instance, if we consider the natural estimator of  $SG_k$ ,

$$\widehat{SG}_k = \frac{1}{\bar{X}} \left( M - (k - 1)m - \int \left( F_n(y) + (k - 1)(1 - F_n(y))^k \right) dy \right)$$

we will derive the following result,

**Theorem 2.** *Let  $(x_1, \dots, x_n)$  a random sample from  $F$  then*

$$\sqrt{n} \frac{\widehat{SG}_k - SG_k}{S_k} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \tag{8}$$

where  $SG_k$  is defined in (1) and,

$$\begin{aligned} S_k^2 = & \frac{1}{\mu^2} \iint_m^y \left( SG_k - 1 + (k - 1)(1 - F(x))^{k-1} \right) \times \\ & \left( SG_k - 1 + (k - 1)(1 - F(y))^{k-1} \right) F(x)(1 - F(y)) dx dy \\ & + \frac{1}{\mu^2} \iint_y^M \left( SG_k - 1 + (k - 1)(1 - F(x))^{k-1} \right) \times \\ & \left( SG_k - 1 + (k - 1)(1 - F(y))^{k-1} \right) F(y)(1 - F(x)) dx dy \tag{9} \end{aligned}$$

**Proof.** Reasoning like in the previous theorem we have the equality

$$\begin{aligned} \sqrt{n}(\widehat{SG}_k - SG_k) = & \frac{\sqrt{n}}{\bar{X}} \left( M - (k - 1)m - \int \left( F_n(y) + (k - 1)(1 - F_n(y))^k \right) dy \right) - \\ & \frac{\sqrt{n} \bar{X}}{\bar{X} \mu} \left( M - (k - 1)M - \int \left( F(y) + (k - 1)(1 - F(y))^k \right) dy \right) \end{aligned}$$

we know that:  $F_n(u) = F(u) + O_{\mathcal{P}}(n^{-1/2})$ , so we can check that,

$$\begin{aligned} & \int \left( F_n(y) + (k - 1)(1 - F_n(y))^k \right) dy \\ & = \int \left( F_n(y) + (k - 1)(1 - F_n(y))^{k-1} (1 - F_n(y)) \right) dy \\ & = \int \left( F(y) + (k - 1)(1 - F(y))^{k-1} (1 - F_n(y)) \right) dy + O_{\mathcal{P}}(n^{-1/2}) \end{aligned}$$

and

$$\begin{aligned}
& \sqrt{n} \left( M - (k-1)m - \int \left( F_n(y) + (k-1)(1-F_n(y))^k \right) dy \right) - \\
& \quad \sqrt{n} \frac{\bar{X}}{\mu} \left( M - (k-1)m - \int \left( F(y) + (k-1)(1-F(y))^k \right) dy \right) \\
& = \sqrt{n} \int \left( F(y) + (k-1)(1-F(y))^k \right) dy - \\
& \quad \sqrt{n} \int \left( F_n(y) + (k-1)(1-F_n(y))^k \right) dy - \\
& \quad \sqrt{n} \left( 1 - \frac{\bar{X}}{\mu} \right) \left( M - (k-1)m - \int \left( F(y) + (k-1)(1-F(y))^k \right) dy \right) \\
& = \int \left( (k-1)(1-F(y))^{k-1} - 1 \right) \sqrt{n} (F_n(y) - F(y)) dy + O_{\mathcal{P}}(n^{-1/2}) - \\
& \quad \frac{\int x d[\sqrt{n}(F(x) - F_n(x))] }{\mu} (M - (k-1)m) - \\
& \quad \frac{\int x d[\sqrt{n}(F(x) - F_n(x))] }{\mu} \int \left( F(y) + (k-1)(1-F(y))^k \right) dy \\
& = \int \left( (k-1)(1-F(y))^{k-1} - 1 + SG_k \right) \sqrt{n} (F_n(y) - F(y)) dy + O_{\mathcal{P}}(n^{-1/2})
\end{aligned}$$

Newly, we apply the Hungarian Construction to obtain that a Brownian Bridge,  $\mathcal{W}^0$ , exists such that we have the equality

$$\begin{aligned}
\Delta_n^k & = \int \left( (k-1)(1-F(y))^{k-1} - 1 + SG_k \right) \sqrt{n} (F_n(y) - F(y)) dy \\
& = \int \left( (k-1)(1-F(y))^{k-1} - 1 + SG_k \right) \mathcal{W}^0\{F(y)\} dy + O_{\mathcal{P}}(n^{-1/2}) \\
& = \Delta^k + O_{\mathcal{P}}(n^{-1/2})
\end{aligned}$$

and from a Brownian Bridge properties and proceeding as in (6) we have that there exists  $\theta \in (m, M)$  such that

$$\begin{aligned}
\Delta^k & = \int \left( (k-1)(1-F(y))^{k-1} - 1 + SG_k \right) \mathcal{W}^0\{F(y)\} dy \\
& = (M - m) \left( SG_k - 1 + k(k-1)(1-F(\theta))^{k-1} \right) \mathcal{W}^0\{F(\theta)\}
\end{aligned}$$

so,  $\Delta^k$  is normal distributed with mean zero and variance

$$\begin{aligned}
 E[(\Delta^k)^2] &= \iint (SG_k - 1 + (k - 1)(1 - F(x))^{k-1}) \times \\
 &\quad (SG_k - 1 + (k - 1)(1 - F(y))^{k-1}) E(\mathcal{W}^0\{F(x)\}\mathcal{W}^0\{F(y)\}) dx dy \\
 &= \iint_m^y (SG_k - 1 + (k - 1)(1 - F(x))^{k-1}) \times \\
 &\quad (SG_k - 1 + (k - 1)(1 - F(y))^{k-1}) F(x)(1 - F(y)) dx dy + \\
 &\quad \iint_y^M (SG_k - 1 + (k - 1)(1 - F(x))^{k-1}) \times \\
 &\quad (SG_k - 1 + (k - 1)(1 - F(y))^{k-1}) F(y)(1 - F(x)) dx dy = (\mu\mathcal{S}_k)^2
 \end{aligned}$$

the result (8) is immediately deduced applying a similar reasoning to the one used in (7). □

Theorem 2 is more general than Theorem 1, of course, expressions in (8) and (9) are generalizations of expressions in (3) and (5) and they are the same for  $k = 2$ .

Following, we will apply the previous method to derive the asymptotic normality for Theil Coefficient (Theil 1967) defined as:

$$\mathcal{T} = \frac{1}{\mu} \int_m^M y \log\left(\frac{y}{\mu}\right) dF(y) = \frac{1}{\mu} \left( M \log\left(\frac{M}{\mu}\right) - \int_m^M \left(\log\left(\frac{y}{\mu}\right) + 1\right) F(y) dy \right)$$

the usual way to estimate the Theil coefficient is to consider the estimator

$$\hat{\mathcal{T}}_n = \frac{1}{n\bar{X}} \sum_{i=1}^n x_i \log\left(\frac{x_i}{\bar{X}}\right)$$

Applying the previously technique we will prove the result

**Theorem 3.** *Let  $(x_1, \dots, x_n)$  a random sample from  $F$  then*

$$\sqrt{n} \frac{\hat{\mathcal{T}}_n - \mathcal{T}}{\mathcal{D}} \xrightarrow{\mathcal{L}}_n \mathcal{N}(0, 1)$$

where

$$\begin{aligned}
 \mathcal{D}^2 &= \frac{1}{\mu^2} \iint_m^y \left(2 - \mathcal{T} - \log\left(\frac{x}{\mu}\right)\right) \left(2 - \mathcal{T} - \log\left(\frac{y}{\mu}\right)\right) F(x)(1 - F(y)) dx dy \\
 &\quad + \frac{1}{\mu^2} \iint_y^M \left(2 - \mathcal{T} - \log\left(\frac{x}{\mu}\right)\right) \left(2 - \mathcal{T} - \log\left(\frac{y}{\mu}\right)\right) F(y)(1 - F(x)) dx dy
 \end{aligned}$$

**Proof.** We have that

$$\begin{aligned}\sqrt{n}(\hat{T} - T) &= \frac{\sqrt{n}}{\bar{X}} \left( M \log \left( \frac{M}{\bar{X}} \right) - \int_m^M \left( \log \left( \frac{x}{\bar{X}} \right) + 1 \right) F_n(x) dx \right) \\ &\quad - \frac{\sqrt{n}}{\bar{X}} \frac{\bar{X}}{\mu} \left( M \log \left( \frac{M}{\mu} \right) - \int_m^M \left( \log \left( \frac{x}{\mu} \right) + 1 \right) F(x) dx \right)\end{aligned}$$

and, if we define the variable,

$$\begin{aligned}\eta_n &= \sqrt{n} \left( M \log \left( \frac{M}{\bar{X}} \right) - \int_m^M \left( \log \left( \frac{x}{\bar{X}} \right) + 1 \right) F_n(x) dx \right) \\ &\quad - \sqrt{n} \frac{\bar{X}}{\mu} \left( M \log \left( \frac{M}{\mu} \right) - \int_m^M \left( \log \left( \frac{x}{\mu} \right) + 1 \right) F(x) dx \right)\end{aligned}$$

we have the equality

$$\begin{aligned}\eta_n &= \sqrt{n} M \left( \log \left( \frac{\mu}{\bar{X}} \right) + \left( 1 - \frac{\bar{X}}{\mu} \right) \log \left( \frac{M}{\mu} \right) \right) - \\ &\quad \sqrt{n} \left( \int_m^M \log \left( \frac{\mu}{\bar{X}} \right) F_n(y) dy - \int_m^M \left( \log \left( \frac{y}{\mu} \right) + 1 \right) (F_n(y) - F(y)) dy \right) \\ &\quad + \sqrt{n} \left( \frac{\bar{X}}{\mu} - 1 \right) \int_m^M \left( \log \left( \frac{y}{\mu} \right) + 1 \right) F(y) dy\end{aligned}$$

Using one-term Taylor expansion for the logarithmic function in a neighborhood of one, we have

$$\sqrt{n} \log \left( \frac{\mu}{\bar{X}} \right) = \sqrt{n} \log \left( 1 + \frac{\mu - \bar{X}}{\bar{X}} \right) = \sqrt{n} \frac{\mu - \bar{X}}{\bar{X}} + O_{\mathcal{P}}(n^{-1/2}) \quad (10)$$

From the equalities  $\bar{X} = \mu + O_{\mathcal{P}}(n^{-1/2})$  and  $F_n(u) = F(u) + O_{\mathcal{P}}(n^{-1/2})$  and (10) we can obtain the equality

$$\begin{aligned}\eta_n &= \frac{M}{\mu} \left( 1 + \log \left( \frac{M}{\mu} \right) \right) \int_m^M \sqrt{n} (F_n(x) - F(x)) dx \\ &\quad - \frac{1}{\mu} \int_m^M F(y) dy \int_m^M \sqrt{n} (F_n(x) - F(x)) dx \\ &\quad - \int_m^M \left( \log \left( \frac{y}{\mu} \right) + 1 \right) \sqrt{n} (F_n(y) - F(y)) dy \\ &\quad - \frac{1}{\mu} \int_m^M \left( \log \left( \frac{y}{\mu} \right) + 1 \right) F(y) dy \int_m^M \sqrt{n} (F_n(x) - F(x)) dx + O_{\mathcal{P}}(n^{-1/2})\end{aligned}$$

we know that  $\int_m^M F(x) dx = M - \mu$  so we have that

$$\eta_n = \int_m^M \left( 2 - T - \log \left( \frac{y}{\mu} \right) \right) \sqrt{n} (F_n(y) - F(y)) dy + O_{\mathcal{P}}(n^{-1/2})$$



and applying the Hungarian Construction we know that a Brownian Bridge,  $\mathcal{W}^0$ , exists such that

$$\eta_n = \int_m^M \left( 2 - \mathcal{T} - \log \left( \frac{y}{\mu} \right) \right) \mathcal{W}^0 \{F(x)\} dx + O_{\mathcal{P}}(n^{-1/2})$$

then, from Integral Mean Value Theorem, for  $\theta \in (m, M)$  we have the equality

$$\begin{aligned} \eta &= \int_m^M \left( 2 - \mathcal{T} - \log \left( \frac{x}{\mu} \right) \right) \mathcal{W}^0 \{F(x)\} dx \\ &= (M - m) \left( 2 - \mathcal{T} - \log \left( \frac{\theta}{\mu} \right) \right) \mathcal{W}^0 \{F(\theta)\} \end{aligned}$$

and  $\eta$  is normal distribution with mean zero and variance

$$\begin{aligned} E(\eta^2) &= E \left( \iint \left( 2 - \mathcal{T} - \log \left( \frac{x}{\mu} \right) \right) \mathcal{W}^0 \{F(x)\} \left( 2 - \mathcal{T} - \log \left( \frac{y}{\mu} \right) \right) \mathcal{W}^0 \{F(y)\} dx dy \right) \\ &= \iint \left( 2 - \mathcal{T} - \log \left( \frac{x}{\mu} \right) \right) \left( 2 - \mathcal{T} - \log \left( \frac{y}{\mu} \right) \right) E[\mathcal{W}^0 \{F(x)\} \mathcal{W}^0 \{F(y)\}] dx dy \\ &= \iint_m^y \left( 2 - \mathcal{T} - \log \left( \frac{x}{\mu} \right) \right) \left( 2 - \mathcal{T} - \log \left( \frac{y}{\mu} \right) \right) F(x)(1 - F(y)) dx dy + \\ &\quad \iint_y^M \left( 2 - \mathcal{T} - \log \left( \frac{x}{\mu} \right) \right) \left( 2 - \mathcal{T} - \log \left( \frac{y}{\mu} \right) \right) F(y)(1 - F(x)) dx dy = (\mu \mathcal{D})^2 \end{aligned}$$

and finally we have

$$\sqrt{n} \frac{\widehat{T}_n - \mathcal{T}}{\mathcal{D}} \stackrel{\text{a.s.}}{=} \frac{\eta_n}{\bar{X} \mathcal{D}} = \frac{\eta}{\mu \mathcal{D}} + O_{\mathcal{P}}(n^{-1/2})$$

and the result is completed. □

In practice, previous theorems can not be used because it is impossible to compute neither the expected value nor variance. In proposition 2.1 of Martínez-Cambor (2006) is proved that if there exists real values  $0 \leq m < M < \infty$  such that  $F(m) = 0$  and  $F(M) = 1$  (this assumption is assumed always in this work);  $F$  has three bounded and continuous derivatives; the kernel function used in (2),  $K$ , has bounded variation; its support is contained in a compact set; and the parameter  $h_n$  satisfies that

$$\frac{\log h_n^{-1}}{nh_n^3} \rightarrow_n 0$$

then, we have the equality

$$\sup_{t \in \mathbb{R}} \left| \widetilde{F}_n(t) - F(t) \right| = o_{\mathcal{P}}(h_n)$$

As consequence of this, if we define

$$\begin{aligned}
 G_n &= \frac{1}{X} \int \tilde{F}_n(u)(1 - \tilde{F}_n(u)) du \\
 \mathcal{V}_n^2 &= \frac{1}{X^2} \iint_m^u (1 - 2\tilde{F}_n(u) + G_n)(1 - 2\tilde{F}_n(v) + G_n) \tilde{F}_n(v)(1 - \tilde{F}_n(u)) dv du \\
 &\quad + \frac{1}{X^2} \iint_u^M (1 - 2\tilde{F}_n(u) + G_n)(1 - 2\tilde{F}_n(v) + G_n) \tilde{F}_n(u)(1 - \tilde{F}_n(v)) dv du \\
 \mathcal{T}_n &= \frac{1}{X} \int u \log\left(\frac{u}{X}\right) d\tilde{F}_n(u) \\
 \mathcal{D}_n^2 &= \frac{1}{X^2} \iint_m^y \left(2 - \mathcal{T}_n - \log\left(\frac{x}{X}\right)\right) \times \\
 &\quad \left(2 - \mathcal{T}_n - \log\left(\frac{y}{X}\right)\right) \tilde{F}_n(x)(1 - \tilde{F}_n(y)) dx dy \\
 &\quad + \frac{1}{X^2} \iint_y^M \left(2 - \mathcal{T}_n - \log\left(\frac{x}{X}\right)\right) \times \\
 &\quad \left(2 - \mathcal{T}_n - \log\left(\frac{y}{X}\right)\right) \tilde{F}_n(y)(1 - \tilde{F}_n(x)) dx dy
 \end{aligned}$$

then, we will have the convergence  $G_n \xrightarrow{\text{a.s.}} G$ ,  $\mathcal{V}_n^2 \xrightarrow{\text{a.s.}} \mathcal{V}^2$  and

$$\sqrt{n} \frac{\hat{G}_n - G_n}{\mathcal{V}_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (11)$$

and,  $\mathcal{T}_n \xrightarrow{\text{a.s.}} \mathcal{T}$ ,  $\mathcal{D}_n^2 \xrightarrow{\text{a.s.}} \mathcal{D}^2$  and

$$\sqrt{n} \frac{\hat{\mathcal{T}}_n - \mathcal{T}_n}{\mathcal{D}_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (12)$$

All these parameters are easily computed. Expressions (11) and (12) can be used to make confidence intervals and to make inferences from a real data set. Obviously, it is straightforward to apply the same method to build confidences intervals for  $SG_k$ .

### 3. Smoothed Bootstrap

In this section we propose a resample method to compute the mean and variance of the studied estimators. When we are assuming that the distribution is continuous the more efficient resample technique is the Smooth Bootstrap. This method is studied, for example, for Hall et al. (1989) or González Manteiga et al. (1994) and if we had a sample of size  $n$  this would be its basic procedure:

1. We compute the SEDF,  $\tilde{F}_n$ , from the data.

2. We run  $B$  bootstrap samples with size  $n$  from  $\tilde{F}_n$  and we compute  $\hat{G}_n$  for each one.
3. We approximate the real distribution of  $\hat{G}_n$  from the distribution of the previous  $B$  computed.

For that estimators, the consistence of the previous method is proved straightforward. For example, for GCI, let  $(x_1^B, \dots, x_n^B)$  be a random sample from  $\tilde{F}_n$  and  $G_n^B$  the Gini coefficient of  $\tilde{F}_n$ , we have the convergence

$$\sqrt{n} \frac{\hat{G}_n^B - G_B}{\mathcal{V}_B} \xrightarrow{\mathcal{L}}_n \mathcal{N}(0, 1)$$

where  $\mu_B = \int x d\tilde{F}_n(x)$ ,

$$\begin{aligned} G_B &= \frac{1}{\mu_B} \int \tilde{F}_n(u)(1 - \tilde{F}_n(u)) du \\ \mathcal{V}_B^2 &= \frac{1}{\mu_B^2} \int \int_m^u (1 - 2\tilde{F}_n(u) + G_B)(1 - 2\tilde{F}_n(v) + G_B) \tilde{F}_B(v)(1 - \tilde{F}_n(u)) dv du \\ &\quad + \frac{1}{\mu_B^2} \int \int_u^M (1 - 2\tilde{F}_n(u) + G_B)(1 - 2\tilde{F}_n(v) + G_B) \tilde{F}_n(u)(1 - \tilde{F}_n(v)) dv du \end{aligned}$$

The same process can be applied without wrinkles to Theil and S-Gini coefficients.

## 4. Simulations

Finally, we describe the performance of both proposed methods in three different distributions and for different sample sizes. In each case we carry out a thousand Monte Carlo (Metropolis & Ulam 1949) samples and we compute confidence intervals for Gini index using the asymptotic distribution given in (3) and when we estimate the mean and variance from the Smoothed Bootstrap method. For this work we have used the software R (R Development Core Team 2006).

In the first situation, we consider a Weibull distribution with parameters three and two. Its real Gini coefficient value is 0.2063 and the standard deviation of Gini estimator (value of the parameter defined in (5)) is  $\mathcal{V} = 0.1490$ .

In table 1 we can be see as both the Gini coefficient and the standard deviation are well approximated but a little overestimate and the asymptotic approximation for intervals works always better than bootstrap percentiles although the coverage probability is lower than the nominal  $1 - \alpha$  level. On the other hand, for sample sizes  $n \geq 100$ , the coverage probability of asymptotic intervals are quite similar.

In the second case (table 2), the distribution considered is Logarithmic Normal with parameters zero and one. The values of the Gini and its deviation  $\mathcal{V}$  are 0.5205 and 0.4271 respectively.

As in the first case, the approximations for the Gini coefficient and its standard deviation are quite good; but there is to much intervals that exclude the real value

TABLE 1: Percentage of miscoverage in tails.

	$n = 50$		$n = 100$		$n = 250$		$n = 500$	
	left	right	left	right	left	right	left	right
95%	7.50	0.40	5.70	0.10	4.00	1.00	3.80	1.90
99%	1.60	0.00	1.00	0.10	1.20	0.10	0.70	0.70
$Mean(\mathcal{V}_n)$	0.1555		0.1525		0.1513		0.1501	
95% (B)	13.30	0.40	11.20	0.10	7.50	0.50	7.00	1.30
99% (B)	4.05	0.00	2.95	0.10	2.90	0.10	1.90	0.60
$Mean(\mathcal{V}_B)$	0.1523		0.1526		0.1505		0.1499	

of the Gini in the left side when we used the bootstrap method and in the right side when asymptotic distribution is used. The obtained percentage is near to the expected one only for  $n = 500$  and we used the asymptotic approximation.

TABLE 2: Percentage of miscoverage in tails.

	$n = 50$		$n = 100$		$n = 250$		$n = 500$	
	left	right	left	right	left	right	left	right
95%	3.80	8.30	4.60	6.80	4.40	4.70	1.60	4.20
99%	2.20	3.30	1.60	2.50	1.50	1.00	0.20	1.00
$Mean(\mathcal{V}_n)$	0.3217		0.3512		0.3704		0.4241	
95% (B)	11.80	3.80	16.40	2.80	12.50	1.90	11.30	1.20
99% (B)	4.40	1.60	7.40	1.40	6.80	0.40	6.00	1.30
$Mean(\mathcal{V}_B)$	0.3920		0.3887		0.3953		0.4068	

Finally, in the third case, we consider the incomes of 5426 Spanish families in 1998 (this data set are from the European Community Household Panel) and we suppose that its Smoothed Empirical Distribution Function (SEDF) is the real distribution function (the Kernel Density Estimation (KDE) and the Smoothed Empirical Distribution Function (SEDF) appear in figure 1). The mean of families incomes is 10298.8 euros with a standard deviation of 7298.82. The Gini index is 0.3531 and the standard deviation of its estimator is 0.2682.

In the table 3 we can see that the  $\mathcal{V}$  estimations are good but both, asymptotic and bootstrap approximations, have a too high number of intervals which don't contain the real value of the parameter and the convergence is slow.

In general the obtained results are the usual in these kind of studies. We obtain quite good fitted for not especially big sample sizes and, of course, with smaller sample sizes than the usual ones in this type of studies.

TABLE 3: Coverage Probability.

	$n = 50$	$n = 100$	$n = 250$	$n = 500$
95%	8.00	7.40	8.60	7.90
99%	2.10	2.20	2.30	1.80
$Mean(\mathcal{V}_n)$	0.2437	0.2473	0.2479	0.2475
95% (B)	11.40	9.00	9.80	9.70
99% (B)	4.20	3.80	3.40	3.10
$Mean(\mathcal{V}_B)$	0.2566	0.2609	0.2637	0.2680

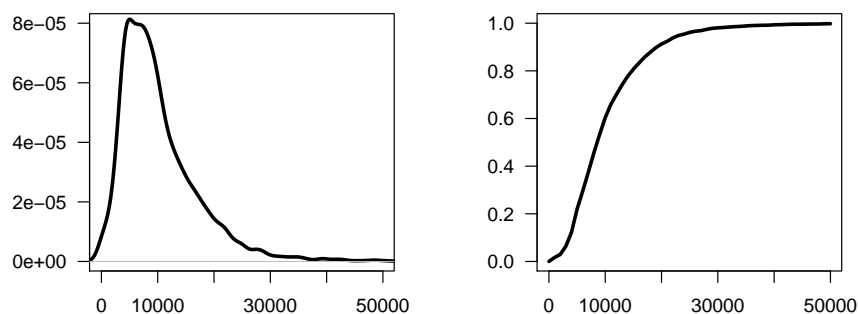


FIGURE 1: KDE and SEDF for Spanish Incomes.

## 5. Conclusions

In this paper, we have not only develop a method for proving the asymptotic normality of Gini, S-Gini and Theil estimators, but we also give speed convergence bounders.

The same method of proof is easily applicable to other indices; for instance it is straightforward to obtain the asymptotic normality of the E-Gini (Chakravarty 1988), a family of coefficients defined to each  $\delta \geq 1$  as:

$$2 \left( \int_0^1 (p - L(p))^\delta dp \right)^{1/\delta}$$

where  $L(p)$  is the Lorenz function.

On the other hand, we have obtained explicit expressions for asymptotic means and variances which are easily estimated and resampling plan has been proposed. Simulations show that asymptotic approximation intervals are always better than bootstrap intervals although the coverage probability is always lower than the nominal  $1 - \alpha$  level.

*Recibido: enero de 2007*  
*Aceptado: agosto de 2007*

## References

- Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley and Sons, New York.
- Chakravarty, S. R. (1988), 'Extended Gini Indices of Inequality', *International Economic Review* **29**, 147–156.
- Crocetta, C. & Loperfido, N. (2005), 'The Exact Sampling Distribution of L-statistics', *Metron* **LXIII**(2), 293.

- Donalson, D. & Weymark, J. A. (1980), 'A Single Parameter Generalization of the Gini Index and Inequality', *Journal of Economic Theory* **22**, 67–86.
- Donalson, D. & Weymark, J. A. (1983), 'Ethical Flexible Indices for Income Distributions in the Continuum', *Journal of Economic Theory* **29**, 353–358.
- Giles, E. (2004), 'Calculating a Standard Error for the Gini Coefficient: Some Further Results', *Oxford Bulletin of Economics & Statistics* **66**, 425–428.
- Gini, C. (1995), Variabilidad e mutabilidad, in E. Pizetti & T. Salvemini, eds, 'Memorie di Metodologia Statistica', Rome. Reprinted (Libreria Eredi Virgilio Veschi, 1912).
- González Manteiga, W., Prada Sánchez, J. M. & Romo, J. (1994), 'The Bootstrap: a Review', *Computational Statistics* **9**, 165–205.
- Hall, P., DiCiccio, J. T. & Romano, J. P. (1989), 'On Smoothing and the Bootstrap', *Annals of Statistics* **17**(2), 692–704.
- Komlós, J., Major, P. & Tusnády, G. (1975), 'An Approximation of Partial Sums of Independent RV-S, and the Sample DF.I.', *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **32**(1-2), 111–131.
- Martínez-Camblor, P. (2005), 'Normalidad asintótica para los E-Gini', *Revista de la Sociedad Argentina de Estadística* **9**, 35–42.
- Martínez-Camblor, P. (2006), 'Tests no paramétricos basados en una medida de igualdad entre funciones de densidad', Servicio de Publicaciones de la Universidad de Oviedo, Oviedo, España.
- Martínez-Camblor, P. (2007), 'Comparing S-Gini and E-Gini on Paired Samples', InterSTAT.  
\*<http://interstat.statjournals.net/YEAR/2007/abstracts/0705001.php>
- Metropolis, N. & Ulam, S. (1949), 'The Monte Carlo Method', *Journal of the American Statistical Association* **44**, 335–345.
- Nadaraya, E. A. (1964), 'Some New Estimates for Distribution Functions', *Theory Prob. Applic.* **9**, 497–500.
- R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
\*<http://www.R-project.org>
- Theil, H. (1967), *Economics and Information Theory*, North-Holland, Amsterdam.
- Yitzhaki, S. (1983), 'On an Extension of the Gini Inequality Index', *International Economic Review* **24**, 183–192.
- Zitikis, R. (2003), 'Asymptotic Estimation of the E-Gini Index', *Econometric Theory* **19**, 587–601.