STATISTICA, anno LXIV, n. 3, 2004

THE VARIANCE OF GINI'S MEAN DIFFERENCE AND ITS ESTIMATORS (*)

M. Zenga, M. Polisicchio, F. Greselin

1. INTRODUCTION

The two most known measures of variability are the standard deviation σ and the Gini's mean difference Δ. The standard deviation σ is more popular than Δ because its square, the variance σ^2, has some interesting properties from both descriptive and sampling points of view. The popularity of the mean difference is mainly due to its relationship with the Gini's concentration index. An impediment to a larger diffusion of Δ is due to some difficulties arising in the determination of the variance of the sample mean difference, $\mathrm{Var}(\hat{\Delta})$, as well as in its estimation.

$\mathrm{Var}(\hat{\Delta})$ results to be a function of σ^2, Δ and a functional $\mathcal{F}$. In this paper a strong relationship between Δ and $\mathcal{F}$ is pointed out: in effect, Δ is the expected value of $D(x)$, the mean deviation about x, and $\mathcal{F}$ is the expected value of $D^2(x)$. Moreover, in this paper two estimators for $\mathrm{Var}(\hat{\Delta})$ are introduced: the natural estimator $\tilde{V}ar(\hat{\Delta})$ and an unbiased estimator $\hat{V}ar(\hat{\Delta})$.

The present work is organized as follows. In Section 2 some definitions and notations are introduced. In Section 3, $\mathrm{Var}(\hat{\Delta})$ is derived for sampling from a continuous random variable (c.r.v.) and for sampling with replacement from a finite population (f.p.).

Section 4 deals with the functional $\mathcal{F}$, its relationship with $D^2(x)$ and its computing formulas. Some examples, referring to four c.r.v. and two finite populations, are also given.

The natural estimator of $\mathrm{Var}(\hat{\Delta})$ is obtained in Section 5, while an unbiased estimator of $\mathrm{Var}(\hat{\Delta})$ is given in Section 6. For the previously selected examples, some sampling simulations are reported in Section 7. Finally, Section 8 concludes and points out some possible developments.

(*) This paper, though it is the result of a close collaboration, was specifically elaborated as follows: Section 6 and the Concluding remarks are due to M. Zenga, M. Polisicchio wrote the Introduction, Sections 2 and 4, while Sections 3, 5 and 7 are due to F. Greselin.

2. NOTATIONS AND DEFINITIONS

Let X be a c.r.v. with probability density function $f(x)$, for $x \in \mathbb{R}$. The Gini's mean difference Δ is defined by:

$$\Delta = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |x-y| f(x) f(y)\, dx\, dy \tag{2.1}$$

Let $D(x)$ be the mean deviation of the c.r.v. X about x:

$$D(x) = \mathrm{E}(|x-X|) = \int_{-\infty}^{+\infty} |x-y| f(y)\, dy \tag{2.2}$$

It is easy to show that:

$$\Delta = \int_{-\infty}^{+\infty} D(x) f(x) dx = \mathrm{E}(D). \tag{2.3}$$

In the case of a f. p. of N units, Δ is given by:

$$\Delta = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} |a_i - a_j|, \tag{2.4}$$

where a_1 , a_2, ..., a_N are the values that the variate X takes on the N units. By considering the mean deviation of X about a_i:

$$D(a_i) = \frac{1}{N}\sum_{j=1}^{N} |a_i - a_j| , \tag{2.5}$$

the mean difference can be expressed by:

$$\Delta = \frac{1}{N}\sum_{i=1}^{N} D(a_i) = \mathrm{M}_1(D). \tag{2.6}$$

Let μ e σ^2 denote, respectively, the mean and the variance of both the variate X of a f.p. and the c.r.v. X. In this paper it is assumed that σ^2 is finite.

By $(X_1,...,X_i,... X_n)$ we denote both a random sample of size n ($n > 3$) from the c.r.v. X and a random sample (with replacement) from the variate X of a f.p.. Obviously, in both cases, the r.v. X_i (i=1,2,...,n) are independent and identically distributed. Let $\hat{\Delta}$ denote the sample mean difference without repetition:

$$\hat{\Delta} = \frac{1}{n(n-1)} \sum_{\substack{i=1,\\ i \neq j}}^{n} \sum_{j=1}^{n} \left| X_i - X_j \right| . \tag{2.7}$$

It is well known that:

$$\mathrm{E}(\hat{\Delta}) = \Delta, \quad \text{for every } \Delta . \tag{2.8}$$

3. THE VARIANCE OF $\hat{\Delta}$

To get the variance of $\hat{\Delta}$ its second moment about zero is needed (see, for example Kendall *et al.*, 1994, p. 362):

$$\begin{aligned} \mathrm{E}(\hat{\Delta}^2) &= \mathrm{E}\left\{ \left[\frac{1}{n(n-1)} \sum_{\substack{i=1\\ i \neq j}}^{n} \sum_{j=1}^{n} \left| X_i - X_j \right| \right]^2 \right\} = \\ &= \frac{1}{n^2 (n-1)^2} \sum_{\substack{i=1\\ i \neq j}}^{n} \sum_{j=1}^{n} \sum_{\substack{s=1\\ s \neq t}}^{n} \sum_{t=1}^{n} \mathrm{E}(\left| X_i - X_j \right| \left| X_s - X_t \right|) . \end{aligned} \tag{3.1}$$

The $n^2(n-1)^2$ terms of the quadruple summation can be classified as follows:

i) $n(n-1)(n-2)(n-3)$ terms with $i \neq j \neq s \neq t$; for these terms:

$$\mathrm{E}(\left| X_i - X_j \right| \left| X_s - X_t \right|) = \mathrm{E}(\left| X_i - X_j \right|) \mathrm{E}(\left| X_s - X_t \right|) = \Delta^2;$$

ii) $n(n-1)$ terms with $i = s, j = t$, $i \neq j$ and $n(n-1)$ terms with $i = t, j = s$, $i \neq j$; for these cases:

$$\mathrm{E}(\left| X_i - X_j \right| \left| X_i - X_j \right|) = \mathrm{E}(\left| X_i - X_j \right| \left| X_j - X_i \right|) = \mathrm{E}(\left| X_i - X_j \right|^2) = 2\sigma^2;$$

iii) $n(n-1)(n-2)$ terms with only two equal indexes of the summation, one in each expression delimited by the absolute value. For these cases:

$$\begin{aligned} \mathrm{E}(\left| X_i - X_j \right| \left| X_i - X_t \right|) &= \mathrm{E}(\left| X_i - X_j \right| \left| X_s - X_i \right|) = \\ &= \mathrm{E}(\left| X_i - X_j \right| \left| X_j - X_t \right|) = \mathrm{E}(\left| X_i - X_j \right| \left| X_s - X_j \right|) = \mathsf{F}; \end{aligned}$$

in conclusion:

$$\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\sum_{\substack{s=1\\ s\neq t}}^{n}\sum_{t=1}^{n}\mathrm{E}(\left|X_i - X_j\right|\left|X_s - X_t\right|) =$$

$$= n(n-1)(n-2)(n-3)\Delta^2 + 4n(n-1)\sigma^2 + 4n(n-1)(n-2)\mathsf{F}.$$

Hence:

$$\mathrm{E}(\hat{\Delta}^2) = \frac{1}{n(n-1)}[4\sigma^2 + 4(n-2)\mathsf{F} + (n-2)(n-3)\Delta^2] \ ,$$

and, finally:

$$\mathrm{Var}(\hat{\Delta}) = \frac{4}{n(n-1)}\left[\sigma^2 + (n-2)\mathsf{F} - \frac{(2n-3)}{2}\Delta^2\right]. \qquad (3.2)$$

Formula (3.2) for $\mathrm{Var}(\hat{\Delta})$ is deduced here as in (Lomnicki, 1952). The variance of $\hat{\Delta}$ was first derived in (Nair, 1936), in a somewhat complex way, on the basis of the ordinal statistics $X_{(i)}$, $i = 1,...,n$. Moreover, Nair derived the expression of the variance of $\hat{\Delta}$ for the Normal, the Exponential and the Rectangular distributions.

Michetti and Dall'Aglio gave an equivalent expression for the variance of $\hat{\Delta}$ and derived it for the Pareto distribution. They also obtained the variance of $\hat{\Delta}$ in the case of unordered sampling without replacement (Michetti and Dall'Aglio, 1957).

Glasser derived the variance of $\hat{\Delta}$ in the case of sampling without replacement (Glasser, 1962). To overcame the issue of the estimators of Δ and their variances, Schechtman and Yitzhaki proposed an upper bound on the variance of $\hat{\Delta}$ (Schechtman and Yitzhaki, 1990).

4. THE FUNCTIONAL $\mathcal{F}$

The functional $\mathcal{F}$ may be expressed in different forms. A simple formula can be based on the mean deviation of X about x, respectively given by (2.2) and (2.5) for c.r.v. and f.p.. This expression allows to obtain an interesting relation between $\mathcal{F}$ and Δ^2.

For a c.r.v., $\mathcal{F}$ is given by:

$$\mathcal{F} = \iiint \left|x - y\right|\left|x - z\right| f(x) f(y) f(z)\, dx\, dy\, dz\,, \qquad (4.1)$$

or:

$$\mathcal{F} = \int\left\{\int|x-y|f(y)dy\right\}\left\{\int|x-z|f(z)dz\right\} f(x)dx\,. \tag{4.2}$$

By identifying the mean deviation of X about x, we get:

$$\mathcal{F} = \int_{-\infty}^{+\infty} D^2(x)f(x)dx = E(D^2). \tag{4.3}$$

and, for (2.3), it follows that:

$$\mathcal{F} \geq \Delta^2. \tag{4.4}$$

In the following, we get an expression for the computation of $\mathcal{F}$, based once again on $D(x)$. In fact:

$$\begin{aligned} D(x) &= \int_{-\infty}^{+\infty} |x-y| f(y)\, dy = \int_{-\infty}^{x} (x-y)f(y)\, dy + \int_{x}^{+\infty} (y-x)f(y)\, dy = \\ &= xF(x) - Q(x) + [\mu - Q(x)] - x[1-F(x)] = \\ &= 2[xF(x) - Q(x)] + (\mu - x) \end{aligned}$$

where $F(x)$ is the cumulative distribution function of X and $Q(x)$ is the first incomplete moment function of X, i.e.: $Q(x) = \int_{-\infty}^{x} y f(y) dy$.

Now:

$$D^2(x) = 4[xF(x) - Q(x)]^2 + (\mu - x)^2 + 4[xF(x) - Q(x)](\mu - x), \tag{4.6}$$

and, by substituting this expression into (4.3), we get:

$$\begin{aligned} \mathcal{F} &= \int\{4[xF(x) - Q(x)]^2 + (\mu - x)^2 + 4[xF(x) - Q(x)](\mu - x)\}f(x)dx = \\ &= \sigma^2 + 4\int\{[xF(x) - Q(x)]^2 + [xF(x) - Q(x)](\mu - x)\}f(x)dx. \end{aligned} \tag{4.7}$$

Lomnicki derived formula (4.7) in another way (Lomnicki, 1952): the present approach, however, making reference to $D(x)$, generally allows an easier calculation of the functional $\mathcal{F}$, as it can be appreciated below in this Section.

For some r.v. it is easier to get $\mathcal{F}$ by the functions:

$$G(x) = 1 - F(x)$$

and:

$$V(x) = \mu - Q(x)\,.$$

With this notation, *D(x)* can be expressed by:

$$D(x) = (x-\mu) - 2[xG(x) - V(x)]$$

and $\mathcal{F}$ can be written as:

$$\mathcal{F} = \sigma^2 + 4\int \{[xG(x) - V(x)]^2 - (x-\mu)[xG(x) - V(x)]\} f(x)dx \tag{4.8}$$

In the case of a f.p.:

$$\begin{aligned}\mathcal{F} &= \frac{1}{N^3}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\left|a_i - a_j\right|\left|a_i - a_l\right| = \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|a_i - a_j\right|\left\{\frac{1}{N}\sum_{l=1}^{N}\left|a_i - a_l\right|\right\} = \\ &= \frac{1}{N}\sum_{i=1}^{N} D(a_i)\left\{\frac{1}{N}\sum_{j=1}^{N}\left|a_i - a_j\right|\right\} = \\ &= \frac{1}{N}\sum_{i=1}^{N} D^2(a_i) = M_1(D^2)\end{aligned} \tag{4.9}$$

If $a_{(1)}, a_{(2)}, \ldots, a_{(N)}$ are the ordered values of $a_1, a_2, \ldots, a_N$, it is easy to show that, analogously to (4.7), $\mathcal{F}$ can be also expressed by:

$$\mathcal{F} = \sigma^2 + \frac{4}{N^3}\sum_{i=1}^{N}(ia_{(i)} - T)^2 + \frac{4}{N^2}\sum_{i=1}^{N}(\mu - a_{(i)})(ia_{(i)} - T) \tag{4.10}$$

where:

$$T_i = \sum_{j=1}^{i} a_{(j)}\,.$$

The value of the variance of $\hat{\Delta}$ is below reported for some c.r.v. and two f.p. For the continuous random variables, the mean deviation *D(x)* is evaluated first, and therefore the mean difference Δ and the functional $\mathcal{F}$ are easily obtained from their relation with the mean deviation. For the finite populations, by straightforward changes to notation, a similar approach is chosen.

4.1. *Normal distribution*

Let X be the normal distribution with probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{for } \sigma > 0.$$

So:

$$D(x) = (x-\mu)\left[2\Phi\left(\frac{x-\mu}{\sigma}\right) - 1\right] + \sigma\sqrt{\frac{2}{\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where $\Phi(x)$ denotes, as usual, the c.d.f. of the standard normal,

$$\Delta = \frac{2\sigma}{\sqrt{\pi}} \quad \text{and} \quad \mathcal{F} = \frac{\sigma^2}{3\pi}(\pi + 6\sqrt{3}),$$

hence:

$$\mathrm{Var}(\hat{\Delta}) = \frac{4\sigma^2}{n(n-1)}\left[\frac{n+1}{3} + \frac{2\sqrt{3}(n-2) - 2(2n-3)}{\pi}\right].$$

4.2. *Exponential distribution*

Let X be the c.r.v. with probability density function:

$$f(x) = \theta\, e^{-\theta x} \quad \text{for} \quad x \geq 0,\ \theta > 0.$$

It is well known that $\sigma^2 = \frac{1}{\theta^2}$ and it is easy to show that:

$$D(x) = x + \frac{2}{\theta} e^{-\theta x} - \frac{1}{\theta}; \quad \Delta = \frac{1}{\theta} \text{ and } \mathcal{F} = \frac{4}{3\theta^2}.$$

Hence:

$$\mathrm{Var}(\hat{\Delta}) = \frac{2(2n-1)}{3n(n-1)} \frac{1}{\theta^2}.$$

4.3. *Rectangular distribution*

Let X be the c.r.v. with probability density function:

$$f(x) = \frac{1}{b-a} \quad \text{for} \quad a \leq x \leq b,\ (b > a).$$

In this case $\sigma^2 = \frac{(b-a)^2}{12}$ and also:

$$D(x) = \frac{(x-a)^2}{b-a} + \frac{b+a}{2} - x; \quad \Delta = \frac{b-a}{3}; \quad \mathcal{F} = \frac{7}{60}(b-a)^2.$$

Hence:

$$\mathrm{Var}(\hat{\Delta}) = \frac{n+3}{45n(n-1)}(b-a)^2.$$

4.4. *Pareto distribution*

Let X be the c.r.v. with probability density function:

$$f(x) = \theta x_0^\theta \, x^{-(\theta+1)} \quad \text{for} \quad x \geq x_0,\ \theta > 2,\ x_0 > 0.$$

In this case $\sigma^2 = \frac{\theta\, x_0^2}{(\theta-1)^2(\theta-2)}$ and:

$$D(x) = x - \frac{\theta x_0}{\theta-1} + \frac{2x}{\theta-1}\left(\frac{x_0}{x}\right)^\theta; \ \Delta = \frac{2\theta\, x_0}{(\theta-1)(2\theta-1)};$$

$$\mathcal{F} = \sigma^2 + \frac{2\theta^2 x_0^2}{(\theta-1)^2(2\theta-1)(3\theta-2)}.$$

Hence:

$$\mathrm{Var}(\hat{\Delta}) = \frac{4}{n(n-1)}\left\{\sigma^2 + (n-2)\left[\sigma^2 + \frac{2\theta^2 x_0^2}{(\theta-1)^2(2\theta-1)(3\theta-2)}\right] - \frac{(2n-3)}{2}\Delta^2\right\}.$$

4.5. *Finite population 1*

i	1	2	3	4	5	6	7
a_i	2	4	5	8	11	15	18

For this population:

$$\mu = \frac{1}{7}\sum_{i=1}^{7} a_i = 9; \quad \sigma^2 = \frac{1}{7}\sum_{i=1}^{7}(a_i - 9)^2 = \frac{212}{7}.$$

Δ and $\mathcal{F}$ can be evaluated as follows. Let:

$$T(a_i) = \sum_{j=1}^{N} \left| a_i - a_j \right| = N \cdot D(a_i).$$

So, from (2.4):

$$\Delta = \frac{1}{N^2} \sum_{i=1}^{N} T(a_i) \tag{4.11}$$

and from (4.9):

$$\mathcal{F} = \frac{1}{N^3} \sum_{i=1}^{N} T^2(a_i). \tag{4.12}$$

Table 1 shows how to evaluate Δ and $\mathcal{F}$ from (4.11) and (4.12).

So, we get:

$$\Delta = \frac{304}{7^2} = 6.204082; \text{ and } \mathcal{F} = \frac{13876}{7^3} = 40.45481.$$

By applying formula (3.2), we finally obtain:

$$\mathrm{Var}(\hat{\Delta}) = \frac{4}{n(n-1)} \left[\frac{212}{7} + (n-2)\,40.45481 - \frac{(2n-3)}{2} 38.49063 \right]. \tag{4.13}$$

TABLE 1

Evaluation of $|a_j - a_i|$ *, Δ and* $\mathcal{F}$ *for the data of population 1*

a_i \ a_j	2	4	5	8	11	15	18	$T(a_i)$	$T^2(a_i)$
2	0	2	3	6	9	13	16	49	2401
4	2	0	1	4	7	11	14	39	1521
5	3	1	0	3	6	10	13	36	1296
8	6	4	3	0	3	7	10	33	1089
11	9	7	6	3	0	4	7	36	1296
15	13	11	10	7	4	0	3	48	2304
18	16	14	13	10	7	3	0	63	3969
								304	13876

4.6. *Finite population 2*

i	1	2	3	4	5
a_i	1	3	5	10	11

This example shows how to get the values of μ, σ^2, Δ and $\mathcal{F}$ by using only one table.

TABLE 2

Evaluation of μ, σ², Δ *and of* $\mathcal{F}$, *by formula (4.10)*

i	$a_{(i)}$	$ia_{(i)}$	T_i	$ia_{(i)}$-T_i	μ-$a_{(i)}$	($ia_{(i)}$-T_i)(μ- $a_{(i)}$)	($ia_{(i)}$-T_i)²	(μ- $a_{(i)}$)²
1	1	1	1	0	5	0	0	25
2	3	6	4	2	3	6	4	9
3	5	15	9	6	1	6	36	1
4	10	40	19	21	-4	-84	441	16
5	11	55	30	25	-5	-125	625	25
	30	117				-197	1106	76

From table 2 we get:

$$\mu = 6; \quad \sigma^2 = \frac{76}{5} = 15.2; \quad \mathcal{F} = 15.2 + \frac{4}{5^3}1106 + \frac{4}{5^2}(-197) = 19.072.$$

It is well known that:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|a_i - a_j\right| = 2\sum_{i=1}^{N} a_{(i)}\{2i - (N+1)\};$$

or:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|a_i - a_j\right| = 4\sum_{i=1}^{N} ia_{(i)} - 2(N+1)\sum_{i=1}^{N} a_{(i)}.$$

Therefore:

$$\Delta = \frac{4\sum_{i=1}^{N} ia_{(i)} - 2(N+1)\sum_{i=1}^{N} a_{(i)}}{N^2}.$$

Hence, from table 2, we get:

$$\Delta = \frac{4(117) - 2(6)(30)}{5^2} = \frac{468 - 360}{25} = 4.32.$$

Finally:

$$\mathrm{Var}(\hat{\Delta}) = \frac{4}{n(n-1)}\left[15.2 + (n-2)19.072 - \frac{(2n-3)}{2}18.6624\right]. \qquad (4.14)$$

5. THE NATURAL ESTIMATOR OF THE VARIANCE OF $\hat{\Delta}$

In section 3 we presented the variance of $\hat{\Delta}$, in the case of sampling with replacement from a f.p. or from a c.r.v. (see 3.2).

The aim of this section is to get the natural estimator of Var($\hat{\Delta}$). It can be easily obtained by substituting, in the expression of Var($\hat{\Delta}$), σ^2, $\mathcal{F}$ and Δ^2, with their respective estimators, derived by considering the sample as a finite population of *n* units.

The variance σ^2 can then be estimated by:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{5.1}$$

where $\bar{X}$ is the sample mean; for the mean difference, the estimator is given by:

$$\tilde{\Delta} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|X_i - X_j\right| \tag{5.2}$$

and, for the functional $\mathcal{F}$:

$$\tilde{\mathcal{F}} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\left|X_i - X_j\right|\left|X_i - X_t\right| \tag{5.3}$$

The natural estimator of Var($\hat{\Delta}$) is hence:

$$\tilde{V}ar(\hat{\Delta}) = \frac{4}{n(n-1)}\left[\tilde{\sigma}^2 + (n-2)\tilde{\mathcal{F}} - \frac{(2n-3)}{2}\tilde{\Delta}^2\right]. \tag{5.4}$$

In order to evaluate the bias of the proposed estimator, the expected values of $\tilde{\sigma}^2$, $\tilde{\mathcal{F}}$ and $\tilde{\Delta}^2$ are needed. It is well known that:

$$\mathrm{E}(\tilde{\sigma}^2) = \frac{n-1}{n}\sigma^2, \tag{5.5}$$

and:

$$\mathrm{E}(\tilde{\Delta}^2) = \mathrm{E}\left[\left(\frac{n-1}{n}\hat{\Delta}\right)^2\right] = \left(\frac{n-1}{n}\right)^2\{\Delta^2 + \mathrm{Var}(\hat{\Delta})\}. \tag{5.6}$$

For the natural estimator of $\mathcal{F}$, given by (5.3), some remarks are needed:

$$\mathrm{E}(\tilde{\mathcal{F}}) = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{n}\mathrm{E}(\left|X_i - X_j\right|\left|X_i - X_t\right|). \tag{5.7}$$

In the summation there are:

i) $n(n-1)(n-2)$ terms with $i \neq j \neq t$; for these terms the expected value is:

$$\mathrm{E}(|X_i - X_j||X_i - X_t|) = \mathcal{F};$$

ii) *n(n-1)* terms with $j = t$ and $i \neq j$; for them the expected value is:

$$\mathrm{E}(|X_i - X_j||X_i - X_j|) = \mathrm{E}(|X_i - X_j|^2) = 2\sigma^2;$$

all other terms have a null expectation.
Hence:

$$\mathrm{E}(\tilde{\mathcal{F}}) = \frac{1}{n^3}\left[n(n-1)(n-2)\mathcal{F} + 2n(n-1)\sigma^2\right]. \tag{5.8}$$

From (5.5), (5.6) and (5.8), we get:

$$\mathrm{E}(\tilde{Var}(\hat{\Delta})) = \mathrm{Var}(\hat{\Delta}) + \left[\frac{4}{n^4(n-1)}\right]\{A(n)\sigma^2 + B(n)\mathcal{F} + C(n)\Delta^2\}, \tag{5.9}$$

where:

$$A(n) = 2n^3 - 11n^2 + 14n - 6;$$
$$B(n) = (n-2)(-7n^2 + 12n - 6);$$
$$C(n) = (2n-3)(6n^2 - 11n + 6)/2.$$

The natural estimator $\tilde{Var}(\hat{\Delta})$ is hence a biased estimator of $\mathrm{Var}(\hat{\Delta})$.

Formula (5.9) also assures that $\tilde{Var}(\hat{\Delta})$ is asymptotically unbiased. Moreover, if the relative bias is defined as:

$$\mathrm{r.b.}\{\tilde{Var}(\hat{\Delta})\} = \frac{\mathrm{E}[\tilde{Var}(\hat{\Delta})] - \mathrm{Var}(\hat{\Delta})}{\mathrm{Var}(\hat{\Delta})},$$

we note that:

$$\lim_{n\to\infty} \mathrm{r.b.}\{\tilde{Var}(\hat{\Delta})\} = 0.$$

This condition is stronger than the asymptotic unbiasedness of $\tilde{Var}(\hat{\Delta})$: it gives informations about the rate of convergence to zero for the bias of $\tilde{Var}(\hat{\Delta})$.

6. UNBIASED ESTIMATOR OF THE VARIANCE OF $\hat{\Delta}$

In order to obtain an unbiased estimator of $\mathrm{Var}(\hat{\Delta})$, unbiased estimators of σ^2 and of $\mathcal{F}$ are first needed. An unbiased estimator of σ^2 is:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2. \tag{6.1}$$

The natural estimator of $\mathcal{F}$ is given by (5.3). From (5.8) we get:

$$\frac{n^3 \mathrm{E}(\tilde{\mathcal{F}}) - 2n(n-1)\sigma^2}{n(n-1)(n-2)} = \mathcal{F}. \tag{6.2}$$

By substituting in (6.2) σ^2 with $\mathrm{E}(S^2)$ and $\tilde{\mathcal{F}}$ with the statistic (5.3), we get:

$$\frac{n^2}{(n-1)(n-2)}\mathrm{E}\left[\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\left|X_i - X_j\right|\left|X_i - X_l\right|\right] - \frac{2\mathrm{E}(S^2)}{(n-2)} = \mathcal{F}$$

and hence:

$$\mathrm{E}\left[\frac{1}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\left|X_i - X_j\right|\left|X_i - X_l\right| - \frac{2S^2}{n-2}\right] = \mathcal{F}.$$

So, an unbiased estimator of $\mathcal{F}$ is given by:

$$\hat{\mathcal{F}} = \frac{1}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\left|X_i - X_j\right|\left|X_i - X_l\right| - \frac{2S^2}{n-2}. \tag{6.3}$$

Now, by substituting in (3.2) σ^2 with S^2, $\mathcal{F}$ with $\hat{\mathcal{F}}$ and Δ^2 with $\hat{\Delta}^2$, we get the statistic:

$$\overset{*}{Var}(\hat{\Delta}) = \frac{4}{n(n-1)}\left[S^2 + (n-2)\hat{\mathcal{F}} - \frac{(2n-3)}{2}\hat{\Delta}^2\right]. \tag{6.4}$$

Now, by remembering that:

$$\mathrm{E}(\hat{\Delta}^2) = [\mathrm{E}(\hat{\Delta})]^2 + \mathrm{Var}(\hat{\Delta}) = \Delta^2 + \mathrm{Var}(\hat{\Delta}),$$

we obtain from (6.4):

$$\begin{aligned}\mathrm{E}\left[\overset{*}{Var}(\hat{\Delta})\right] &= \frac{4}{n(n-1)}\left[\sigma^2 + (n-2)\mathcal{F} - \frac{(2n-3)}{2}(\Delta^2 + \mathrm{Var}(\hat{\Delta}))\right] = \\ &= \frac{4}{n(n-1)}\left[\sigma^2 + (n-2)\mathcal{F} - \frac{(2n-3)}{2}\Delta^2\right] - \frac{2(2n-3)}{n(n-1)}\mathrm{Var}(\hat{\Delta}).\end{aligned}$$

From (3.2), it follows that:

$$\begin{aligned} \mathrm{E}\left[\overset{*}{\hat{V}ar}(\hat{\Delta})\right] &= \mathrm{Var}(\hat{\Delta}) - \frac{2(2n-3)}{n(n-1)}\mathrm{Var}(\hat{\Delta}) = \\ &= \mathrm{Var}(\hat{\Delta})\ \frac{(n-2)(n-3)}{n(n-1)}; \end{aligned} \tag{6.5}$$

and:

$$\mathrm{E}\left[\overset{*}{\hat{V}ar}(\hat{\Delta})\ \frac{n(n-1)}{(n-2)(n-3)}\right] = \mathrm{Var}(\hat{\Delta}). \tag{6.6}$$

From (6.6) and (6.4) an unbiased estimator of Var($\hat{\Delta}$) is hence given by:

$$\hat{V}ar(\hat{\Delta}) = \frac{4}{(n-2)(n-3)}\left[S^2 + (n-2)\hat{\mathcal{F}} - \frac{(2n-3)}{2}\hat{\Delta}^2\right]. \tag{6.7}$$

7. SAMPLING RESULTS

The theoretical results of the previous sections are very relevant for applications. The aim of this section is to observe the behaviour of the estimators $\tilde{V}ar(\hat{\Delta})$ and $\hat{V}ar(\hat{\Delta})$, for increasing values of the sample size n and for different underlying distributions, by sampling methods[1].

The adopted methodology can be described by the following steps:

a) choice of the c.r.v. and of the values of its parameters, evaluation of its descriptive indexes, such as μ, σ^2, Δ, the functional $\mathcal{F}$ and Var($\hat{\Delta}$);
b) choice of the number B of independent pseudorandom samples to be drawn from the c.r.v.;
c) setting of the sample sizes n;
d) simulation of B samples of size n from the c.r.v.;
e) for each sample obtained in d), evaluation of $\hat{\Delta}$, $\tilde{V}ar(\hat{\Delta})$ and $\hat{V}ar(\hat{\Delta})$;
f) for each group of B samples, evaluation of:
   - the mean value $M_1(\hat{\Delta})$, to be compared with Δ;
   - the variance $V(\hat{\Delta})$, to be compared with Var($\hat{\Delta}$).
   - $M_1(\tilde{V}ar(\hat{\Delta}))$ and $M_1(\hat{V}ar(\hat{\Delta}))$, to be compared with Var($\hat{\Delta}$).

As far as step a) is concerned, the families of distributions presented in Section 4 will be considered.

Note that the choice of the parameters is crucial: we need the existence of the first and the second moment of each c.r.v. and this requirement will be accomplished in all examined cases.

---

[1] The authors wish to thank Dr. Paolo Radaelli for his valuable help in processing data for an earlier draft of this paper.

The choice of *B*, the number of samples to be drawn from each c.r.v. can be reasonably fixed by *B* = 5000. Finally, the increasing values 6, 30, 60, 120, and 240 were chosen for the sample size *n*.

7.1. *Normal distribution, with* $\mu = 0$, $\sigma = 5$ (Table 3).

For this distribution $\Delta = 5.641896$ and $\mathcal{F} = 35.899778$.

TABLE 3

Simulation data for the estimation of $\mathrm{Var}(\hat{\Delta})$ *by* $\tilde{V}ar(\hat{\Delta})$ *and* $\hat{V}ar(\hat{\Delta})$, *sampling from the normal distribution (B=5000 samples)*

n	$M_1(\hat{\Delta})$	$V(\hat{\Delta})$	$M_1(\tilde{V}ar(\hat{\Delta}))$	$M_1(\hat{V}ar(\hat{\Delta}))$	$\mathrm{Var}(\hat{\Delta})$
6	5.635187	3.297402	2.458101	3.435431	3.381288
30	5.624399	0.566456	0.522389	0.567180	0.565566
60	5.635056	0.271621	0.264487	0.275657	0.276920
120	5.646730	0.135127	0.133863	0.136679	0.137031
240	5.643642	0.065605	0.067517	0.068231	0.068163

We observe that $M_1(\hat{\Delta})$ is very close to the value of Δ, for the unbiasedness of $\hat{\Delta}$. The values of $V(\hat{\Delta})$, $M_1(\tilde{V}ar(\hat{\Delta}))$ and $M_1(\hat{V}ar(\hat{\Delta}))$ steadily decrease as *n* increases, as $\mathrm{Var}(\hat{\Delta})$ does.

The natural estimator $\tilde{V}ar(\hat{\Delta})$ offers a poor estimation of $\mathrm{Var}(\hat{\Delta})$ for small sample sizes (but it shows a better performance as *n* increases), while the unbiased estimator $\hat{V}ar(\hat{\Delta})$ performs well for all values of *n*.

The last column in table 3 provides the theoretical value of $\mathrm{Var}(\hat{\Delta})$, evaluated by (3.2), as a function of the sample size.

7.2. *Exponential distribution, with* $\theta = 0.2$ (Table 4).

For this distribution, $\sigma^2=25$, $\Delta=5$ and $\mathcal{F}=33.333333$.

TABLE 4

Simulation data for the estimation of $\mathrm{Var}(\hat{\Delta})$ *by* $\tilde{V}ar(\hat{\Delta})$ *and* $\hat{V}ar(\hat{\Delta})$, *sampling from the exponential distribution (B=5000 samples)*

n	$M_1(\hat{\Delta})$	$V(\hat{\Delta})$	$M_1(\tilde{V}ar(\hat{\Delta}))$	$M_1(\hat{V}ar(\hat{\Delta}))$	$\mathrm{Var}(\hat{\Delta})$
6	4.994363	6.031127	3.404964	6.179060	6.111111
30	4.982594	1.131140	0.972381	1.108940	1.130268
60	4.989829	0.555146	0.520780	0.556491	0.560264
120	5.002536	0.279077	0.269150	0.278240	0.278945
240	5.005114	0.135643	0.137380	0.139687	0.139179

7.3. *Rectangular distribution, with parameters a=0, b=1* (Table 5).

In this case, $\sigma^2 = 0.083333$, $\Delta = 0.333333$ and $\mathcal{F} = 0.116667$.

TABLE 5

Simulation data for the estimation of $\mathrm{Var}(\hat{\Delta})$ *by* $\tilde{V}ar(\hat{\Delta})$ *and* $\hat{V}ar(\hat{\Delta})$, *sampling from the rectangular distribution (B=5000 samples)*

n	$M_1(\hat{\Delta})$	$V(\hat{\Delta})$	$M_1(\tilde{V}ar(\hat{\Delta}))$	$M_1(\hat{V}ar(\hat{\Delta}))$	$\mathrm{Var}(\hat{\Delta})$
6	0.333384	0.006514	0.007144	0.006814	0.006667
30	0.332434	0.000842	0.000907	0.000848	0.000843
60	0.333204	0.000387	0.000412	0.000395	0.000395
120	0.333599	0.000189	0.000196	0.000191	0.000191
240	0.333396	0.000091	0.000095	0.000094	0.000094

7.4. *Pareto distribution, first case:* $x_0=2$, $\theta = 4$ (Table 6).

For this c.r.v., $\sigma^2 = 0.88889$, $\Delta = 0.76190$ and $\mathcal{F} = 1.09206$.

TABLE 6

Simulation data for the estimation of $\mathrm{Var}(\hat{\Delta})$ *by* $\tilde{V}ar(\hat{\Delta})$ *and* $\hat{V}ar(\hat{\Delta})$, *sampling from the Pareto distribution with* $x_0=2$, $\theta = 4$ *(B=5000 samples)*

n	$M_1(\hat{\Delta})$	$V(\hat{\Delta})$	$M_1(\tilde{V}ar(\hat{\Delta}))$	$M_1(\hat{V}ar(\hat{\Delta}))$	$\mathrm{Var}(\hat{\Delta})$
6	0.757956	0.317598	0.152838	0.323718	0.352653
30	0.760406	0.066182	0.058303	0.068167	0.068609
60	0.763971	0.034881	0.032720	0.035377	0.034203
120	0.762148	0.016666	0.015852	0.016478	0.017077
240	0.761713	0.008636	0.008493	0.008660	0.008532

7.5. *Pareto distribution, second case:* $x_0 = 1$, $\theta = 3$ (Table 7).

For this distribution, $\sigma^2 = 0.75$, $\Delta = 0.6$ and $\mathcal{F} = 0.87857$.

TABLE 7

Simulation data for the estimation of $\mathrm{Var}(\hat{\Delta})$ *by* $\tilde{V}ar(\hat{\Delta})$ *and* $\hat{V}ar(\hat{\Delta})$, *sampling from the Pareto distribution with* $x_0 = 1$, $\theta = 3$ *(B=5000 samples)*

n	$M_1(\hat{\Delta})$	$V(\hat{\Delta})$	$M_1(\tilde{V}ar(\hat{\Delta}))$	$M_1(\hat{V}ar(\hat{\Delta}))$	$\mathrm{Var}(\hat{\Delta})$
6	0.595250	0.278668	0.128952	0.283815	0.352571
30	0.598957	0.062708	0.055119	0.064760	0.069379
60	0.603939	0.034635	0.032320	0.035029	0.034630
120	0.599629	0.015848	0.015113	0.015729	0.017300
240	0.600024	0.009020	0.008805	0.008983	0.008646

For the last two cases, let us remark that the values chosen for the shape parameter θ ($\theta = 3$ and $\theta = 4$) assure the existence of the second moment, but not that of the fourth moment.

All data obtained from the simulations show that $V(\hat{\Delta})$, $\tilde{V}ar(\hat{\Delta})$ and $\hat{V}ar(\hat{\Delta})$ decrease as n increases, like $\mathrm{Var}(\hat{\Delta})$. $V(\hat{\Delta})$ and the unbiased estimator $\hat{V}ar(\hat{\Delta})$ give a good approximation of $\mathrm{Var}(\hat{\Delta})$ for any value of the sample size n (except the second case of the Pareto distribution, for which it holds only for $n \geq 30$) while $\tilde{V}ar(\hat{\Delta})$, being only asymptotically unbiased, improves its estimate as n increases.

In particular, for all considered distributions, the minimum sample sizes that assures a mean relative error $\dfrac{\tilde{V}ar(\hat{\Delta}) - Var(\hat{\Delta})}{Var(\hat{\Delta})} < 0.02$ in estimating $\mathrm{Var}(\hat{\Delta})$ by $\tilde{V}ar(\hat{\Delta})$ is $n = 240$.

In the reminder of this Section, we deal with the two f.p. introduced in Section 3. Now, as sampling with replacement is considered, the whole sample space can be generated, and this is computationally feasible in a reasonable time for low values of n. This allows to verify the unbiasedness of $\hat{V}ar(\hat{\Delta})$ and to evaluate the bias of $\tilde{V}ar(\hat{\Delta})$, even if no new informations will arise.

7.6. *Finite population 1* (Table 8).

i	1	2	3	4	5	6	7
a_i	2	4	5	8	11	15	18

For this population $\mu = 9; \sigma^2 = \dfrac{212}{7}; \Delta = 6.204082$ and $\mathsf{F} = 40.45481$. The sample space has 7^n elements: the exact distribution of the statistics $\hat{\Delta}$, $\tilde{V}ar(\hat{\Delta})$ and $\hat{V}ar(\hat{\Delta})$ can be obtained, as well as their mean values. The results, shown in table 8, have to be exactly matched with the corresponding values obtained by (5.9) and (3.2). The following remarks hold:

- the value of $M_1(\hat{\Delta})$ expresses the unbiasedness of $\hat{\Delta}$,
- its variance $V(\hat{\Delta})$ is the quantity estimated by $\tilde{V}ar(\hat{\Delta})$ and $\hat{V}ar(\hat{\Delta})$,
- $M_1(\tilde{V}ar(\hat{\Delta}))$ precisely matches with (5.9),
- $M_1(\hat{V}ar(\hat{\Delta}))$, evaluated by the exact distribution of $\hat{V}ar(\hat{\Delta})$ on the sample space, agrees with (4.13).

TABLE 8

Data for the estimation of $\hat{\Delta}$ *and* $\mathrm{Var}(\hat{\Delta})$ *for f.p.1, obtained by generation of the whole sample space, compared with the theoretical values of* $\mathrm{E}(\tilde{V}ar(\hat{\Delta}))$ *and* $\mathrm{Var}(\hat{\Delta})$

n	$M_1(\hat{\Delta})$	$V(\hat{\Delta})$	$M_1(\tilde{V}ar(\hat{\Delta}))$	$M_1(\hat{V}ar(\hat{\Delta}))$	$E(\tilde{V}ar(\hat{\Delta}))$	$\mathrm{Var}(\hat{\Delta})$
4	6.204082	4.989588	4.875208	4.989588	4.875208	4.989588
6	6.204082	2.519617	2.750860	2.519617	2.750860	2.519617
8	6.204082	1.630392	1.848289	1.630392	1.848289	1.630392
10	6.204082	1.189060	1.364161	1.189060	1.364161	1.189060

7.7. *Finite population 2* (Table 9).

i	1	2	3	4	5
a_i	1	3	5	10	11

For this population, $\mu = 6$, $\sigma^2 = 15.2$, $\Delta = 4.32$ and $\mathsf{F} = 19.072$.
By generating the whole sample space, the results in Table 9 were obtained.

TABLE 9

Data for the estimation of $\hat{\Delta}$ *and* $\mathrm{Var}(\hat{\Delta})$ *for f.p.2, obtained by generation of the whole sample space, compared with the theoretical values of* $\mathrm{E}(\tilde{V}ar(\hat{\Delta}))$ *and* $\mathrm{Var}(\hat{\Delta})$

n	$M_1(\hat{\Delta})$	$V(\hat{\Delta})$	$M_1(\tilde{V}ar(\hat{\Delta}))$	$M_1(\hat{V}ar(\hat{\Delta}))$	$E(\tilde{V}ar(\hat{\Delta}))$	$\mathrm{Var}(\hat{\Delta})$
4	4.32	2.229333	2.575000	2.229333	2.575000	2.229333
6	4.32	1.000960	1.398637	1.000960	1.398637	1.000960
8	4.32	0.594743	0.893687	0.594743	0.893687	0.594743
10	4.32	0.406471	0.628157	0.406471	0.628157	0.406471
12	4.32	0.301964	0.470407	0.301964	0.470407	0.301964

The simulated data and the exact distributions presented in this section verify all theoretical results above introduced, about the properties of the natural estimator $\tilde{V}ar(\hat{\Delta})$ and the unbiased estimator $\hat{V}ar(\hat{\Delta})$ of $\mathrm{Var}(\hat{\Delta})$.

8. CONCLUDING REMARKS

The variance of the sample Gini's mean difference $\hat{\Delta}$ was already known in the literature (Nair, 1936; Lomnicki, 1952; Michetti and Dall'Aglio, 1957): from (3.2), $\mathrm{Var}(\hat{\Delta})$ can be written as follows:

$$\mathrm{Var}(\hat{\Delta}) = \frac{1}{n} \cdot 4 \cdot \left[\frac{\sigma^2}{(n-1)} + \frac{(n-2)}{(n-1)} \mathcal{F} - \frac{(n-1.5)}{(n-1)} \Delta^2 \right]. \tag{8.1}$$

In this work a useful relationship that ties the definition of the functional $\mathcal{F}$ and the mean difference Δ to *D(x)*, the mean deviation of the r.v. X about *x,* is shown. The functional $\mathcal{F}$ is actually the expected value of $D^2(x)$, while Δ is the expected value of $D(x)$. Some examples for well known continuous random variables allow to appreciate how this approach generally simplify the evaluation of $\mathcal{F}$ and hence $\text{Var}(\hat{\Delta})$ can be obtained more easily.

The natural estimator of $\text{Var}(\hat{\Delta})$ is given by:

$$\tilde{V}ar(\hat{\Delta}) = \frac{1}{n} \cdot 4 \cdot \left[\frac{\tilde{\sigma}^2}{(n-1)} + \frac{(n-2)}{(n-1)} \tilde{\mathsf{F}} - \frac{(n-1.5)}{(n-1)} \tilde{\Delta}^2 \right], \tag{8.2}$$

where:

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \text{ , } \tilde{\Delta} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| X_i - X_j \right|,$$

and:

$$\tilde{\mathcal{F}} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \left| X_i - X_j \right| \left| X_i - X_l \right|.$$

$\tilde{V}ar(\hat{\Delta})$ is an asymptotically unbiased estimator of $\text{Var}(\hat{\Delta})$. Moreover, the relative bias is such that:

$$\lim_{n \to \infty} \frac{\text{E}(\tilde{V}ar(\hat{\Delta})) - \text{Var}(\hat{\Delta})}{\text{Var}(\hat{\Delta})} = 0 .$$

We show that an unbiased estimator of $\text{Var}(\hat{\Delta})$ is given by:

$$\hat{V}ar(\hat{\Delta}) = \frac{1}{(n-2)} \cdot 4 \cdot \left[\frac{S^2}{(n-3)} + \frac{(n-2)}{(n-3)} \hat{\mathcal{F}} - \frac{(n-1.5)}{(n-3)} \hat{\Delta}^2 \right], \tag{8.3}$$

where S^2 and $\hat{\mathcal{F}}$ are the unbiased estimators of σ^2 and $\mathcal{F}$, respectively given by (6.1) and (6.3). Sampling simulations from four c.r.v. allow to better inspect the behaviour of $\tilde{V}ar(\hat{\Delta})$ for increasing values of the sample size n; they also confirmed all theoretical results shown in this work about the properties of $\hat{V}ar(\hat{\Delta})$.

A first conclusion about the estimator $\hat{\Delta}$ of Δ is based on the following relations:

i) $\text{E}(\hat{\Delta}) = \Delta$;

ii) $\text{Var}(\hat{\Delta}) = \frac{1}{n} \cdot 4 \cdot \left[\frac{\sigma^2}{(n-1)} + \frac{(n-2)}{(n-1)} \mathcal{F} - \frac{(n-1.5)}{(n-1)} \Delta^2 \right] \to 0, \quad \text{as } n \to \infty.$

Hence $\hat{\Delta}$ is a mean-squared-error consistent estimator of Δ.

We now consider the standardized sample statistic:

$$T_1 = \frac{(\hat{\Delta}-\Delta)}{\sqrt{\mathrm{Var}(\hat{\Delta})}} = \frac{\sqrt{n}(\hat{\Delta}-\Delta)}{2\sqrt{\frac{\sigma^2}{(n-1)} + \frac{(n-2)}{(n-1)}\mathcal{F} - \frac{(n-1.5)}{(n-1)}\Delta^2}}. \tag{8.4}$$

As $\hat{\Delta}$ is asymptotically normally distributed [Hoeffding, 1948; David, 1981, p. 273], the asymptotic distribution of T_1 is standard normal.

To make inferences regarding Δ, T_1 needs to be modified. For example, to obtain confidence intervals for Δ it is necessary to substitute in (8.4) $\mathrm{Var}(\hat{\Delta})$ with $\hat{V}ar(\hat{\Delta})$.

So we get the sample statistic:

$$T_2 = \frac{(\hat{\Delta}-\Delta)}{\sqrt{\hat{V}ar(\hat{\Delta})}} = \frac{\sqrt{n-2}(\hat{\Delta}-\Delta)}{2\sqrt{\frac{S^2}{(n-3)} + \frac{(n-2)}{(n-3)}\hat{\mathcal{F}} - \frac{(n-1.5)}{(n-3)}\hat{\Delta}^2}}. \tag{8.5}$$

Other analyses are needed to specify the minimum value of n needed to approximate the distribution of T_2 with that of the standard normal.

We recall that the estimator S^2 of σ^2 is such that:

i) $E(S^2) = \sigma^2$;

ii) $\mathrm{Var}(S^2) = \frac{1}{n}\left\{\mu_4 - \frac{(n-3)}{(n-1)}(\sigma^2)^2\right\}$,

where μ_4 is the fourth central moment of the r.v. X [Mood *et al.,* 1974, p. 229]. Hence, S^2 is a mean-squared-error consistent estimator of σ^2.

Note that for the existence of $\mathrm{Var}(S^2)$ it is necessary that $\mu_4 < \infty$, while for the existence of $\mathrm{Var}(\hat{\Delta})$ it is sufficient that $\sigma^2 < \infty$.

Dipartimento di Metodi Quantitativi per le Scienze Economiche e Aziendali Università Milano-Bicocca

MICHELE ZENGA
MARCELLA POLISICCHIO
FRANCESCA GRESELIN

REFERENCES

H.A. DAVID (1981), *Order statistics,* John Wiley and Sons, New York.

G.J. GLASSER (1962), *Variance formulas for the mean difference and coefficient of concentration*, "Journal of the American Statistical Association*",* Vol. 57, pp. 648-654.

W. HOEFFDING (1948), *A Class of Statistics with Asymptotically Normal Distribution*, "The Annals of Mathematical Statistics", Vol. 19, pp. 293-325.

M.G. KENDALL, A. STUART, J.K. ORD (1994), *Distribution theory,* Halsted press, New York.

Z.A. LOMNICKI (1952), *The standard error of Gini's mean difference*, "The Annals of Mathematical Statistics", Vol. 23, pp. 635-637.

B. MICHETTI and G. DALL'AGLIO (1957), *La differenza semplice media,* "Statistica", 17, pp. 159-254.

A.M. MOOD, F.A. GRAYBILL and D.C. BOES (1974), *Introduction to the theory of statistics*, McGraw-Hill, New York.

U.S. NAIR (1936), *The standard error of Gini's mean difference*, "Biometrika", Vol. 28, pp. 428-436.

E. SCHECHTMAN and S. YITZHAKI (1990), *On an upper bound on the variance of Gini's mean difference,* "Metron", Vol. 48, pp. 321-332.

RIASSUNTO

Varianza della differenza media campionaria di Gini e suoi stimatori

L'utilizzo della differenza media di Gini, come indice di variabilità, è stato finora limitato a causa delle difficoltà che si incontrano nella determinazione della varianza della differenza media campionaria $\hat{\Delta}$, così come nella sua stima. Questo lavoro è un contributo all'affronto di tali problematiche. Considerando lo scarto medio assoluto $D(x)$ di una v.c. X rispetto ad un valore assegnato x, si ricava che la differenza media campionaria Δ è il valore atteso di $D(x)$. Inoltre, denotando con $\mathcal{F}$ il valore atteso di $D^2(x)$, Var($\hat{\Delta}$) può essere espresso come funzione di σ^2 (varianza di X), di Δ e $\mathcal{F}$.

Successivamente, sono stati ricavati due stimatori di Var($\hat{\Delta}$): a partire dallo stimatore naturale, di cui si è dimostrata la correttezza asintotica, si è poi ricavato uno stimatore non distorto.

SUMMARY

The variance of Gini's mean difference and its estimators

The use of Gini's mean difference as an index of variability has, until now, been restricted because of some difficulties arising in computing and estimating the variance of its estimator $\hat{\Delta}$. The aim of this paper is to cope with these issues. Considering the mean deviation $D(x)$ of a r.v. X about a given value x, the Gini's mean difference Δ results to be the expected value of $D(x)$. Moreover, denoting by $\mathcal{F}$ the expected value of $D^2(x)$ and by $\hat{\Delta}$ the sample mean difference without repetition, Var($\hat{\Delta}$) can be expressed as a function of the variance of X, say σ^2, Δ and $\mathcal{F}$.

Two estimators for Var($\hat{\Delta}$) are obtained: starting from the natural estimator, whose asympthotic unbiasedness is shown, an unbiased estimator is then derived.