GINI'S MEAN DIFFERENCE IN THE THEORY AND APPLICATION TO INFLATED DISTRIBUTIONS

T. Gerstenkorn, J. Gerstenkorn

1. INTRODUCTION

In 1911 Prof. Corrado Gini published (in Italian) a very vast statistical study initiating consideration on the mean called later in the literature Gini's mean difference. Curiously enough, the subsequent (non-Italian) authors dealing with this problem do not refer to that work. In the book by Kendall and Stuart (1963), Gini's name was mentioned and his work was inserted in the references, but without further particulars. Therefore, it may be supported that, for many authors, the work in question was difficult of attainment. It is written in an ancient style: very lengthily (156 pages) with long descriptions. It is hard to perceive any modern notation in it.

The period of World War I undoubtedly disturbed the extension of Gini's ideas. In the twenties C. Gini referred to his idea. In this case he published two papers in foreign journals (1921, 1926). We have failed to ascertain whether someone was dealing with the mean difference in the thirties and forties. Maybe, the language barrier of the published papers caused not widespread popularity among not-Italian theoreticians of statistics. It was only in fifties and further decades when the Italian statisticians discussed anew (in Italian) the mean difference. We see here Salvemini (1956 and 1957), Michetti and Dall'Aglio (1957), Castellano (1965), Girone (1968a, 1968b), Zanardi (1973 and 1974). However, it is worthy of notice since, unlike other quantities designed for measuring the dispersion of a random variable, the mean difference is independent of any central measure of localization, which can be seen from its definition

$$\Delta_1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |x - y|\, dF(x)dF(y)\,. \tag{1.1}$$

When the random variable X is discrete (a case more often considered) the formula has the form

$$\Delta_1 = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} |x_i - x_j| \, p_i p_j \,, \tag{1.2}$$

where $p_i = P(X = x_i)$, $p_j = P(X = x_j)$.

The analytic investigation of the discussed characteristic is made difficult because of the absolute value occurring in the formula. However, it facilitates the computations on numerical data, which also concerns, as is well known, the mean deviation. Hence we sometimes encounter the investigations concerning the mean difference, connected with the mean deviation. This is the case, for instance, in Ramasubban (1958). Methodically, this paper is based on operations indicated by Johnson (1957) in the considerations referring to the mean deviation of the binomial distribution.

The mean deviation, also for the binomial distribution, was considered earlier by Frame (1945).

The difficulties connected with the absolute value can efficiently be overcome in the case of the mean deviation by using incomplete moments. This was demonstrated in T. Gerstenkorn's paper (1975).

As far as the mean difference is concerned, the investigation of this statistic and, in particular, of some of its properties referring to a random sample did not give any adequate results although, in the case of a normal variable, one can mention a few papers. For the normal distribution, the exact standard error of the mean difference was probably given for the first time by Nair (1936), but with the application of a rather complicated method. Much later, in 1952, Lomnicki obtained the very result by using a simpler method. A year later, Kamat calculated the third moment in the exact form and, in the considerations on the skewness measure β_1, inferred that, for a great n, the distribution of the mean difference may be the same as the χ-distribution. Following Kamat, Ramasubban (1956) obtained an approximation of values for the fourth moment and showed that the concentration measure (kurtosis) β_2 calculated on this basis, taken together with the values for β_1 (obtained by Kamat), seems to show the exactness of the χ-approximations, at least for sample greater than 10 ($n > 10$). The same author tried to obtain an empirical distribution of the mean difference for small samples ($n < 10$), but we do not know whether the results were published.

The normal distribution may be considered as the limit case of the binomial and the Poisson one under certain assumptions. So, it is natural to examine the moments from the sample Δ_1 for those discrete distributions and to try to adapt a suitable distribution for Δ_1, even if an exact distribution cannot be found easily. In order to make the task easier, at the first step one derives the formulae for the absolute mean difference Δ_r given as

$$\Delta_r = \sum_i \sum_{j \neq i} |x_i - x_j|^r \, p_i p_j \,. \tag{1.3}$$

This problem is discussed in Ramasubban's paper (1959). An extension of the problem can be found in the paper by Katti (1960).

The work of Gini is also mentioned by Rao (1982).

Gini's mean difference met with no approbation of the authors of handbooks. We did not find it in any Polish handbook. It is concisely discussed in the English handbook by Kendall and Stuart (1963). It is worth our while to mention here the German textbook by Rinne (1974) where a practical application of this statistic was discussed.

2. PROPERTIES OF THE MEAN DIFFERENCE

If a random variable takes a finite number of values then the expression for the difference is written down in the form

$$\Delta_1 = \frac{1}{N^2}\sum_{i=1}^{l}\sum_{j=1}^{l}|x_i - x_j|\, n_i n_j\,, \tag{2.1}$$

where $\frac{n_i}{N} \approx p_i = P(X = x_i)$, $\frac{n_j}{N} \approx p_j = P(X = x_j)$ and $n_1 + n_2 + ... + n_l = N$, and it is the so-called difference with repetitions.

The difference is sometimes defined in another way

$$\Delta_1 = \frac{1}{N(N-1)}\sum_{i=1}^{l}\sum_{j=1}^{l}|x_i - x_j|\, n_i n_j, \qquad i \neq j \tag{2.2}$$

as the mean difference without repetitions.

Sometimes, the above formulae are written down without taking the weights into account, and then

$$\Delta_1 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|x_i - x_j|,\ i, j = 1,2,...,N\,, \tag{2.1a}$$

$$\Delta_1 = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}|x_i - x_j|, \qquad i \neq j\,. \tag{2.2a}$$

However, we most often use the following formulae

$$\Delta_1 = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}|x_i - x_j| \tag{2.2b}$$

or

$$\Delta_1 = \frac{2}{N(N-1)} \sum_{j=1}^{N} \sum_{i=1}^{j-1} |x_i - x_j| . \tag{2.2c}$$

These formulae will be illustrated by an example from Rinne's book (p. 119)

$$\sum_{\nu=1}^{n} \sum_{\substack{\kappa=1 \\ \kappa>\nu}}^{n} |x_\nu - x_\kappa|$$

TABLE 1

x_o \ x_κ	1	5	6	6	8	10	13	13	16	22	$\sum_{\kappa=\nu+1}^{n} \lvert x_\nu - x_\kappa \rvert$
1	–	4	5	5	7	9	12	12	15	21	90
5	–	–	1	1	3	5	8	8	11	17	54
6	–	–	–	0	2	4	7	7	10	16	46
6	–	–	–	–	2	4	7	7	10	16	46
8	–	–	–	–	–	2	5	5	8	14	34
10	–	–	–	–	–	–	3	3	6	12	24
13	–	–	–	–	–	–	–	0	3	9	12
13	–	–	–	–	–	–	–	–	3	9	12
16	–	–	–	–	–	–	–	–	–	6	6
22	–	–	–	–	–	–	–	–	–	–	–
$\sum_{\nu=1}^{\kappa-1} \lvert x_\nu - x_\kappa \rvert$	–	4	6	6	14	24	42	42	66	120	324

From Table 1 we get

$$\Delta_1 = \frac{2}{10 \cdot 9} 324 = 7,2 .$$

It might seem that the difficulties with the occurrence of the absolute value will disappear if, in place of the mean difference, we introduce the coefficient

$$E^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-y)^2 dF(x) dF(y) .$$

However, after simple calculations it turns out that

$$E^2 = 2\mu_2 ,$$

is a double variance. Nevertheless, this interesting relation shows that the variance may be defined as a half of the mean value of the squares of all possible differences of values of the variable, that is, in other words, one can define it with no need of turning to a consideration of deviations with respect to the central (mean) value. We found this remark in Udny Yule and Kendall (1953) (p. 146) as in Kendall and Stuart only (p. 47; in Russ. ed. p. 74).

The notation of formulae (1.1) for the mean difference may be so modified that there will not be the absolute value sign. Note that

$$\sum_{i=1}^{l}\sum_{j=1}^{l}|x_i - x_j|\,n_i n_j = 2\sum_{i=j+1}^{l}\sum_{j=1}^{l-1}|x_i - x_j|\,n_i n_j\,. \tag{2.3}$$

If the observations are marked with numbers in such a way that

$$x_1 < x_2 < \ldots < x_N\,,$$

then formula (2.3) may be written down in the form

$$2\sum_{i=j+1}^{l}\sum_{j=1}^{l-1}(x_i - x_j)n_i n_j\,, \tag{2.3a}$$

and then formulae (2.1) and (2.2) will take the form

$$\Delta_1 = \frac{2}{N^2}\sum_{i=j+1}^{l}\sum_{j=1}^{l-1}(x_i - x_j)n_i n_j\,, \tag{2.4}$$

$$\Delta_1 = \frac{2}{N(N-1)}\sum_{i=j+1}^{l}\sum_{j=1}^{l-1}(x_i - x_j)n_i n_j\,. \tag{2.5}$$

Formulae (2.4) and (2.5) may be given some other form. Note that, after careful calculations, we have

$$\sum_{i=j+1}^{N}\sum_{j=1}^{N-1}(x_i - x_j) = \sum_{k=1}^{N-1}k(N-k)(x_{k+1} - x_k),$$

therefore the mean difference may be written down as

$$\Delta_1 = \frac{2}{N^2}\sum_{k=1}^{N-1}k(N-k)(x_{k+1} - x_k) \tag{2.6}$$

or

$$\Delta_1 = \frac{2}{N(N-1)}\sum_{k=1}^{N-1}k(N-k)(x_{k+1} - x_k)\,. \tag{2.7}$$

These forms of the mean are particularly handy when the distances $x_{k+1} - x_k$ are the same.

A further simplification of the formulae can be obtained by introducing a distribution function

$$F_k = P(X \le x_k) = F(x_k).$$

In the case when $F_k = \frac{k}{N}$ and the distances are identical and equal to unity, we get

$$\Delta_1 = \frac{2}{N^2} \sum_{k=1}^{N-1} NF_k(N - NF_k) = 2 \sum_{k=1}^{N-1} F_k(1 - F_k). \tag{2.8}$$

If we denote by $G_k = NF_k$ the cumulated frequency, then we shall obtain

$$\Delta_1 = \frac{2}{N^2} \sum_{k=1}^{N-1} G_k(N - G_k). \tag{2.9}$$

This form is convenient for practical computations, which is demonstrated by the Table (Kendall and Stuart, pp. 50-51; Russ. ed. p. 78)

TABLE 2

Height, inches	Frequency	G_k	$N - G_k$	$G_k \ N - G_k$
57–	2	2	8583	17,166
58–	4	6	8579	51,474
59–	14	20	8565	171,300
60–	41	61	8524	519,964
61–	83	144	8441	1,215,504
62–	169	313	8272	2,589,136
63–	394	707	7878	5,569,746
64–	669	1376	7209	9,919,584
65–	990	2366	6219	14,714,154
66–	1223	3589	4996	17,930,644
67–	1329	4918	3667	18,034,306
68–	1230	6148	2437	14,982,676
69–	1063	7211	1374	9,907,914
70–	646	7857	728	5,719,896
71–	392	8249	336	2,771,664
72–	202	8451	134	1,132,434
73–	79	8530	55	469,150
74–	32	8562	23	196,926
75–	16	8578	7	60,046
76–	5	8583	2	17,166
77–	2	8585	–	–
Totals	8585	–	–	105,990,850

From Table 2 we obtain

$$\Delta_1 = \frac{2 \cdot 105990850}{8585^2} = 2{,}88.$$

To make the presentation full, it is worth our while to mention the so-called concentration coefficient of Gini. Gini was engaged in the question of concentration as early as 1910, but he gave it a proper form in the paper of 1914 presented on the 29th of May (*i.e.* shortly before the outbreak of World War I) at the meeting of the Royal Venetia Institute for Science, Letters and the Arts:

$$G = \frac{\Delta_1}{2m}, \qquad m = E(X)\text{, if it exists,}$$

or

$$G = \frac{\Delta_1}{2\overline{x}}$$

which is, of course, an abstract number.

In statistical practice we also use of the so-called concentration curve of Lorenz (1905). It is a curve whose points have the co-ordinates $(F(x), \Phi(x))$ where

$$\Phi(x) = \frac{1}{m}\int_{-\infty}^{x} x dF(x)$$

is the so-called incomplete moment $(0 \le \Phi(x) \le 1)$. The curve is convex. It can be shown that the area *S*, contained between the concentration curve and the straight line $\Phi = F$, is equal numerically to $\frac{1}{2}G$. The proof can be found in Kendall and Stuart (p. 49; Russ. ed. pp. 76-77).

3. THE MEAN DIFFERENCE FOR INFLATED DISTRIBUTIONS

In many fields of science, one applies the well-known probability distributions of a discrete random variable. However, there happen situations in which we are ready to admit that a given phenomenon is subject to a typical probability distribution, under the condition that we shall expose this distribution to some deformation. In such a case, we apply most frequently the so-called mixture of distributions. As the simplest mixture we may classify the so-called *inflated distribution* which consists in composing any discrete distribution with the degenerate (i.e. one-point) distribution.

We shall introduce the following notations for the discrete distribution:

$$P(X = i) = h(i), \quad i = 0,1,2,...$$

We then have

Definition 3.1. We say that a discrete random variable *Y* is subject to the inflated distribution (deformed at the point $i = 0$) if its probability function is expressed by the formula

$$P(Y = i) = \begin{cases} \beta + \alpha\, h(0) & \text{if} \quad i = 0, \\ \alpha\, h(i) & \text{if} \quad i > 1, \end{cases} \tag{3.1}$$

where $\alpha \in (0,1]$, while $\beta = 1 - \alpha$.

The deformation of a distribution may also take place at any point of the distribution.

Definition 3.2. We say that a discrete random variable Y is subject to the *generalized inflated distribution* (i.e. the one with a deformation at any point $i = l$) if

$$P(Y=i) = \begin{cases} \beta + \alpha\, h(l) & \text{if} \quad i = l, \\ \alpha\, h(i) & \text{if} \quad i = 0,1,2,...,l-1,l+1,..., \end{cases} \tag{3.2}$$

where $\alpha \in (0,1]$ and $\beta = 1 - \alpha$.

In particular, for the binomial distribution, we have

Definition 3.3. We say that a discrete random variable Y is subject to the inflated binomial distribution $P(X=i)$ (deformed at the point $i = 0$) if its probability function is expressed by the formula

$$P(Y=i) = \begin{cases} \beta + \alpha\, q^n & \text{if} \quad i = 0, \\ \alpha \binom{n}{i} p^i q^{n-i} & \text{if} \quad i = 1, 2, ..., n, \end{cases} \tag{3.3}$$

where $\alpha \in (0,1]$, while $\beta = 1 - \alpha$, $0 < p < 1$, $p + q = 1$.

If $\alpha = 1$, then the above distribution reduces to the binomial distribution

$$P(X=i) = \binom{n}{i} p^i q^{n-i} \quad \text{for} \quad i = 0,1,2,...,n.$$

Definition 3.4. We say that a discrete random variable Y is subject to the generalized inflated binomial distribution if its probability function is expressed by the formula

$$P(Y=i) = \begin{cases} \beta + \alpha \binom{n}{l} p^l q^{n-l} & \text{if} \quad i = l, \\ \alpha \binom{n}{i} p^i q^{n-i} & \text{if} \quad i = 0,1,2,...,l-1,l+1,...,n, \end{cases} \tag{3.4}$$

where $0 < \alpha \le 1$, $\alpha + \beta = 1$, $0 < p < 1$, $p + q = 1$.

Of course, formulae (3.1) – (3.4) present probability distributions, which follows from simple calculations.

Inflated distributions were introduced into the literature by S. N. Singh (1963) for the case of the Poisson distribution, and next made a thorough study of for

the binomial distribution by M. P. Singh for deformations at the initial point (1965/66) and at an arbitrary one (1966).

Inflated distributions were being dealt with by many authors. Many papers on various subjects were written. The problems concerning these distributions were discussed, for instance, by T. Gerstenkorn (1977).

The mean differences for distribution (3.1) and for inflated binomial distribution (3.3) were discussed by T. Gerstenkorn in the paper of 1997.

Here we shall deal with the mean difference for generalized inflated discrete distribution (3.2).

Theorem 3.1. Gini's mean difference for the generalized inflated discrete distribution is expressed by the formula

$$\Delta_1 = 2\alpha\beta\{[2lF(l+1)-1] - m_1 + 2\,m_1(l+1)\} \tag{3.5}$$

$$+ 2\alpha^2[\sum_{j=1}\sum_{i=0}^{j-1}(j-i)h(i)h(j) - \sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)],$$

where:

m_1 – the expected value of an uninflated distribution,

$m_1(l+1)$ – the right-hand incomplete moment (i.e. the one with the truncation of the value of the variable to $x = l$ inclusive) of the uninflated distribution,

$F(l+1)$ – the distribution function of the uninflated distribution at a point $x = l+1$.

Proof. From the definition we have

$$\Delta_1 = \sum_{i=0}^{j-1}\sum_{j=1}(j-i)P(Y=i)P(Y=j) + \sum_{i=1}\sum_{j=0}^{i-1}(i-j)P(Y=i)P(Y=j)$$

$$= \sum_{j=l+1}\sum_{\substack{i=0\\ i\neq l}}^{j-1}(j-i)P(Y=i)P(Y=j) + lP(Y=0)P(Y=l) + (l-1)P(Y=1)P(Y=l) + \ldots$$

$$+ P(Y=l-1)P(Y=l) + \sum_{j=l+1}(j-l)P(Y=l)P(Y=j)$$

$$+ \sum_{i=l+1}\sum_{\substack{j=0\\ j\neq l}}^{i-1}(i-j)P(Y=j)P(Y=i) + lP(Y=0)P(Y=l) + (l-1)P(Y=1)P(Y=l) + \ldots$$

$$+ P(Y=1)P(Y=l) + \sum_{i=l+1}(i-l)P(Y=l)P(Y=i) =$$

$$= \alpha^2 \sum_{j=l+1} \sum_{\substack{i=0 \\ i\neq l}}^{j-1} (j-i)h(i)h(j) + \alpha l h(0)(\beta + \alpha h(l))$$

$$+ \alpha(l-1)h(1)(\beta + \alpha h(l)) + ... + \alpha h(l-1)(\beta + \alpha h(l))$$

$$+ \alpha \sum_{j=l+1} (j-l)(\beta + \alpha h(l)h(j)) + \alpha^2 \sum_{i=l+1} \sum_{\substack{j=0 \\ j\neq l}}^{i-1} (i-j)h(i)h(j)$$

$$+ \alpha l h(0)(\beta + \alpha h(l)) + \alpha(l-1)h(1)(\beta + \alpha h(l)) + ...$$

$$+ \alpha h(l-1)(\beta + \alpha h(l)) + \alpha \sum_{i=l+1} (i-l)(\beta + \alpha h(l)h(i)$$

$$= 2\alpha^2 \sum_{j=l+1} \sum_{\substack{i=0 \\ i\neq l}}^{j-1} (j-i)h(i)h(j) + 2\alpha\beta l h(0) + 2\alpha^2 l h(0)h(l))$$

$$+ 2\alpha\beta(l-1)h(1) + 2\alpha^2(l-1)h(1)h(l) + ... + 2\alpha\beta h(l-1)$$

$$+ 2\alpha^2 h(l-1)h(l) + 2\alpha\beta \sum_{j=l+1} (j-l)h(j) + 2\alpha^2 \sum_{j=l+1} (j-l)h(l)h(j)$$

$$= 2\alpha\beta[l h(0) + (l-1)h((1) + ... + h(l-1)] + 2\alpha\beta \sum_{j=l+1} (j-l)h(j)$$

$$+2\alpha^2 \sum_{j=l} \sum_{i=0}^{j-1} (j-i)h(i)h(j)$$

$$= 2\alpha\beta \sum_{j=0}^{l-1} (l-j)h(j) + 2\alpha\beta \sum_{j=l+1} (j-l)h(j)$$

$$+2\alpha^2 \sum_{j=1} \sum_{i=0}^{j-1} (j-i)h(i)h(j) - 2\alpha^2 \sum_{j=1}^{l-1} \sum_{i=0}^{j-1} (j-i)h(i)h(j)$$

$$= 2\alpha\beta\, l \sum_{j=0}^{l} h(j) - 2\alpha\beta \sum_{j=0}^{l} jh(j) + 2\alpha\beta \sum_{j=l+1} jh(j)$$

$$- 2\alpha\beta\, l \sum_{j=l+1} h(j) + 2\alpha^2 \sum_{j=1} \sum_{i=0}^{j-1} (j-i)h(i)h(j) - 2\alpha^2 \sum_{j=1}^{l-1} \sum_{i=0}^{j-1} (j-i)h(i)h(j)$$

$$= 2\alpha\beta\, l\left[F(l+1)-(1-\sum_{j=0}^{l}h(j)\right]-2\alpha\beta\left(\sum_{j=0}jh(j)-\sum_{j=l+1}jh(j)\right)$$

$$+2\alpha\beta\sum_{j=l+1}jh(j)+2\alpha^2\sum_{j=1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)$$

$$= 2\alpha\beta\, l[F(l+1)-(1-F(l+1)]-2\alpha\beta[m_1-m_1(l+1)]$$

$$+2\alpha\beta m_1(l+1)+2\alpha^2\sum_{j=1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j),$$

which already implies formula (3.5).

Basing ourselves on (3.5), we shall demonstrate what forms the mean difference takes for generalized inflated binomial distribution (3.4). For the purpose, we shall make use of relation (2.4), p. 550 from Ramasubban's paper (1958) as well as formula (1.14) for the incomplete moment of the binomial distribution, cited in T. Gerstenkorn's paper (1971) as the result given by Risser and Traynard (1933, pp. 320-321 or 1957, pp. 92-93).

Namely,

$$m_1(l+1)=(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}+npm_0(l+1)$$

where $m_0(l+1)=\sum_{i=l+1}^{n}P(X=i)=1-F(l+1).$

Taking the above into account, we shall get

$$\Delta_1=4\alpha\beta\, lF(l+1)-2\alpha\beta\, l-2\alpha\beta\, np+4\alpha\beta\left[(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}+npm_0(l+1)\right]$$

$$+2\alpha^2 npq\left[\sum_{i=1}^{n-1}\binom{n-1}{i-1}\binom{n-1}{i}p^{2i-1}q^{2n-2i-1}+\sum_{i=0}^{n-1}\binom{n-1}{i}^2p^{2i}q^{2n-2i-2}\right]$$

$$-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)$$

$$= 4\alpha\beta lF(l+1)-2\alpha\beta l-2\alpha\beta np+4\alpha\beta(l+1)\binom{n}{l+1}\ p^{l+1}q^{n-l}$$

$$+4\alpha\beta npm_0(l+1)+2\alpha^2 npq\left[\sum_{i=1}^{n-1}\binom{n-1}{i-1}\binom{n-1}{i}p^{2i-1}q^{2n-2i-1}+\sum_{i=0}^{n-1}\binom{n-1}{i}^2 p^{2i}q^{2n-2i-2}\right]$$

$$-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j).$$

Finally we obtain

Corollary 3.1. Gini's mean difference in the case of the generalized inflated binomial distribution is expressed by the formula

$$\Delta_1 = 2\alpha\beta\left[2lF(l+1)-l+np(2m_0(l+1)-1)+2(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}\right]$$

$$+2\alpha^2 npq\left[\sum_{i=1}^{n-1}\binom{n-1}{i-1}\binom{n-1}{i}p^{2i-1}q^{2n-2i-1}+\sum_{i=0}^{n-1}\binom{n-1}{i}^2 p^{2i}q^{2n-2i-2}\right]$$

$$-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j). \tag{3.6}$$

The value of the distribution function $F(l+1)$ of the binomial distribution, occurring in (3.6), can be read in the available statistical table of, for example, Zieliński (1972, p. 150).

One can also obtain another form of this relation by using formula (2.8), p. 550, from the paper by Ramasubban:

$$\Delta_1 = 4\alpha\beta lF(l+1)-2\alpha\beta l-2\alpha\beta np+4\alpha\beta(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}$$

$$+4\alpha\beta npm_0(l+1)+2\alpha^2 pq\sum_{i=0}^{n-1}(-1)^i\binom{n}{i+1}\binom{2i}{i}p^i q^i-2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)$$

whence the implication of

Corollary 3.2. Gini's mean difference for the generalized inflated binomial distribution is expressed by the formula

$$\Delta_1 = 2\alpha\beta\left[2lF(l+1)-l+np(2m_0(l+1)-1)+2(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}\right]$$

$$+2\alpha^2\left[pq\sum_{i=0}^{n-1}(-1)^i\binom{n}{i+1}\binom{2i}{i}p^i q^i-\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)\right]. \tag{3.7}$$

One can show (after rather toilsome calculations) that the mean difference $\Delta_1^{'}$ for the uninflated binomial distribution may be written down in the form

$$\Delta_1^{'} = 2pq\sum_{i=0}^{n-1}(-1)^i \binom{n}{i+1}\binom{2i}{i} p^i q^i$$

$$= 2npq\left[1 - \frac{n-1}{1!}\left(\frac{1}{2}\cdot\frac{1}{2}\right)\cdot\frac{4pq}{1!} + \frac{(n-1)(n-2)}{2!}\cdot\frac{1}{3}\left(\frac{1}{2}\cdot\frac{3}{2}\right)\cdot\frac{(4pq)^2}{2!}\right.$$

$$\left. - \ldots + (-1)^{n-1}\frac{(n+1)\cdot\ldots\cdot(2n-2)}{(n-1)!}\cdot\frac{1}{n}\cdot\left(\frac{1}{n-1}\cdot\frac{n}{n-1}\right)\cdot\frac{(4pq)^{n-1}}{(n-1)!}\right].$$

If we adopt the notations

$$a = -(n-1), \quad b = \frac{1}{2}, \quad c = 2, \quad x = 4pq,$$

then $$\Delta_1^{'} = 2npq\left[1 + \frac{ab}{c}\cdot\frac{x}{1!} + \frac{a(a+1)b(b+1)}{c(c+1)}\cdot\frac{x^2}{2!} + \ldots\right.$$

$$\left. + \frac{a(a+1)\ldots(a+n-2)\cdot b(b+1)\ldots(b+n-2)}{c(c+1)\ldots(c+n-2)}\cdot\frac{x^{n-1}}{(n-1)!}\right],$$

which can be written down in a simpler way as

$$\Delta_1^{'} \cong 2npqF(a,b,c,x) = 2npqF\left[(-n+1), \frac{1}{2}, 2, 4pq\right], \tag{3.8}$$

that is, in the form of the hypergeometric series

$$F(a,b,c,x) = 1 + \sum_{n=1}^{\infty}\frac{a^{[n,-1]}b^{[n,-1]}}{c^{[n,-1]}}\cdot\frac{x^n}{n!},$$

where $a^{[n-1]} = a(a+1)(a+2)\ldots(a+n-1)$ is the so-called factorial polynomial (the generalized power).

Taking account of (3.8), we finally obtain

Corollary 3.3. Gini's mean difference of the generalized inflated binomial distribution is expressed by the asymptotic formula

$$\Delta_1 \cong 2\alpha\beta\left[2lF(l+1)-l+np(2m_0(l+1)-1)+2(l+1)\binom{n}{l+1}p^{l+1}q^{n-l}\right]$$

$$+2\alpha^2\left\{npqF\left[(-n+1),\frac{1}{2},2,4pq\right]-\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)\right\}. \tag{3.9}$$

In the case $i=l=0$, the formulae given here are considerably simplified. As has been mentioned, their full forms can be found in T. Gerstenkorn's paper (1997).

Definition 3.5. We say that a discrete random variable Y is subject to the generalized inflated negative binomial distribution if its probability function is expressed by the formula

$$P(Y=l)=\begin{cases}\beta+\alpha q^{-n}\binom{n+l-1}{l}\left(\frac{p}{q}\right)^l & \text{if } i=l\\ \alpha q^{-n}\binom{n+l-1}{l}\left(\frac{p}{q}\right)^l & \text{if } i=0,1,2,...l-1,l+1,...\end{cases} \tag{3.10}$$

where $0<a\leq 1$, $a+\beta=1$, $0<p<1$, $q-p=1$.

T. Gerstenkorn (1997) has shown that the mean difference of an inflated distribution is given by

$$\Delta_1 = 2a\beta m + 2a^2\sum_{i=1}^{i-1}\sum_{j=0}(i-j)h(i)h(j) \tag{3.11}$$

where m is the expected value of the distribution considered without inflation. Ramasubban (1958) has given formulae ((2.11) and (2.12)) for Gini's mean difference of the negative binomial distribution. By using these relations we obtain

Corollary 3.4. Gini's mean difference for the inflated (i=0) negative binomial distribution is expressed by the formula

$$\Delta_1 = 2a\beta np + 2a^2\, npq\sum_{i=0}^{\infty}(-1)^i\binom{n+i}{i}p^i q^i\,\frac{(2i)!}{i!(i+1)!} \tag{3.12}$$

or

$$\Delta_1 \cong 2a\beta np + 2a^2\, npq\, F(n+1,\tfrac{1}{2},2,-4pq), \tag{3.13}$$

where F, as previously, is a notation of a hypergeometric series $F(a, \beta, \gamma, x)$ with parameters

$$a = n+1,\ \beta = \tfrac{1}{2},\ \gamma = 2,\ x = -4pq.$$

In the case when a deformation of the negative binomial distribution takes place in point $i=l$, we make use of (3.5) and formula (1.21) for an incomplete moment of that distribution given by T. Gerstenkorn (1971):

$$m_1(l+1) = (l+1)\binom{-n}{l+1}(-1)^{l+1}p^{l+1}q^{n-l} + np.$$

After suitable calculations we get then

Corollary 3.5. Gini's mean difference for the generalized inflated negative binomial distribution is given by the formula

$$\Delta_1 = 2\alpha\beta\left\{2lF(l+1) - l + np(2m_0(l+1)-1) + 2(l+1)(-1)^{l+1}\binom{-n}{l+1}p^{l+1}q^{n-l}\right\}$$

$$+2\alpha^2 npq\sum_{i=0}^{\infty}(-1)^i\binom{n+1}{i}\frac{(2i)!}{i!(i+1)!}p^i q^i - 2\alpha^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j). \qquad (3.14)$$

or

$$\Delta_1 \cong 2\alpha\beta\{2lF(l+1) - l + np(2m_0(l+1)-1) + 2(l+1)(-1)^{l+1}\binom{-n}{l+1}p^{l+1}q^{n-l}\}$$

$$+2\alpha^2 npqF(n+1,1/2,-4pq). \qquad (3.15)$$

Definition 3.6. We say that a random variable Y is a subject to the generalized inflated Poisson distribution if its probability function is given by the formula

$$P(Y=i)=\begin{cases}\beta + \alpha e^{-\lambda}\dfrac{\lambda^l}{l!} & \text{dla } i = l\\ \alpha \mathrm{e}^{-\lambda}\dfrac{\lambda^{\mathrm{i}}}{i!} & \text{dla } i = 0,1,...,l-1,l+1,...\end{cases} \qquad (3.16)$$

where $0<a\leq 1$, $a+\beta = 1$, $\lambda > 0$.

Making use of (3.11) and of formulae (2.14), (2.15), (2.18)-(2.20) given by Ramasubban (1958), we get

Corollary 3.6. Gini's mean difference of inflated (i=0) Poisson distribution is given by

$$\Delta_1=2a\beta\lambda + 2a^2\lambda e^{-2\lambda}\left[\sum_{i=0}^{\infty}\frac{\lambda^{2i}}{i!i!}+\sum_{i=0}^{\infty}\frac{\lambda^{2i+1}}{i!(i+1)!}\right] \tag{3.17}$$

or replacing these sums by modified Bessel function of the first kind, we get

$$\Delta_1=2a\beta\lambda+ 2a^2\lambda e^{-\lambda}[I_0(2\lambda) + I_1(2\lambda)] \tag{3.18}$$

or also

$$\Delta_1=2a\beta\lambda+ 2a^2 \int_0^{\lambda} e^{-2\lambda} I_0(2\lambda)d\lambda \tag{3.19}$$

or in the form

$$\Delta_1 \cong 2a\beta\lambda+ 2a^2\lambda F(\tfrac{1}{2}, 2, -4\lambda). \tag{3.20}$$

In the case when $i=l$ we make use of (3.5) and of formula (1.16) by T. Gerstenkorn (1971)

$$m_1(l+1) = \lambda[\frac{e^{-\lambda}\lambda^l}{l!}+1-F(l+1)].$$

After suitable calculations , we get then

Corollary 3.7. Gini's mean difference of the generalized inflated Poisson distribution is given by

$$\Delta_1=2a\beta[2F(l+1)(l-\lambda)-l+\lambda+2\frac{e^{-\lambda}\lambda^{l+1}}{l!}]+2a^2\lambda e^{-\lambda}[I_0(2\lambda)+I_1(2\lambda)]-2a^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j) \tag{3.21}$$

or in the form

$$\Delta_1 \cong 2a\beta[2F(l+1)(l-\lambda)-l+\lambda+2\frac{e^{-\lambda}\lambda^{l+1}}{l!}]+2a^2\lambda F(\tfrac{1}{2}, 2, -4\lambda)-2a^2\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j). \tag{3.22}$$

Definition 3.7. We say that a random variable Y is a subject to the generalized inflated logarithmic distribution if its probability function is given by the formula

$$P(Y=i) = \begin{cases} \beta+\alpha c\dfrac{p^l}{l} & \text{if } i=l \\ \alpha c\dfrac{p^i}{i} & \text{if } i=1,2,...l-1,l+1,... \end{cases} \tag{3.23}$$

where $c = -\frac{1}{\ln(1-p)}$, $0<p<1$.

Following as above, we get

Corollary 3.7. Gini's mean difference for the inflated (i=0) logarithmic distribution is given by

$$\Delta_1 = 2a\beta m - 2a^2 \frac{\ln\left[(1-p^2)(1+p)^p\right]}{(1-p)\ln^2(1-p)}, \tag{3.24}$$

where $m=-\frac{p}{(1-p)\ln(1-p)}$ (see: T. Gerstenkorn (1971), formula (2.13)).

In the case when $i=l$ we make use of (3.5) and of formula (2.26) by Ramasubban (1958) and also of formula

$$m_1(l+1) = -\frac{p^{l+1}}{(1-p)\ln(1-p)}$$ (see: T. Gerstenkorn (1971), formula (2.12)).

Then, Gini's mean difference for the generalized inflated logarithmic distribution is given by

$$\begin{aligned}\Delta_1 = 2a\beta[2lF(l+1) - l + \frac{p}{(1-p)\ln(1-p)} - 2\frac{p^{l+1}}{(1-p)\ln(1-p)}] \\ -2a^2 \frac{\ln\left[(1-p^2)(1+p)^p\right]}{(1-p)\ln^2(1-p)} + \sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j).\end{aligned} \tag{3.25}$$

Definition 3.8. We say that a random variable is a subject to the generalized inflated geometric distribution if its probability function is given by

$$P(Y=i) = \begin{cases} \beta + \alpha q p^l & \text{if} \quad i = l \\ \alpha p^i q & \text{if} \quad i = 0,1,...,l-1,l+1,... \end{cases} \tag{3.26}$$

where $a+\beta=1$, $p+q=1$, $p>0$, $q>0$, $\beta\geq 0$, $a>0$.

Making use of (3.11) and (2.28) by Ramasubban (1958), we get

Corollary 3.9. Gini's mean difference for the inflated (i=0) geometric distribution is given by

$$\Delta_1 = 2a\beta\frac{p}{q} + 2\alpha^2\frac{p}{1-p^2}, \tag{3.27}$$

where $m=\frac{p}{q}$.

In the case when $i=l$ we make use of (3.5) and of formula for the incomplete moment of this distribution

$$m_1(l+1) = (l+1)p^{l+1} + \frac{p}{q}(1-F(l+1))$$ (see: T. Gerstenkorn (1971), formula (2.7).

We then have

Corollary 3.10. Gini's mean difference for the generalized inflated geometric distribution is given by

$$\Delta_1=2a\beta[2lF(l+1)-l-2\frac{p}{q}+2\frac{p}{q}(1-F(l+1))+2(l+1)p^{l+1}]$$

$$-2\alpha^2[\frac{p}{1-p^2}+\sum_{j=1}^{l-1}\sum_{i=0}^{j-1}(j-i)h(i)h(j)]. \qquad (3.28)$$

University of Trade TADEUSZ GERSTENKORN
Faculty of Mathematics, University of Łódź

Chair of Statistical Methods, Institute of Econometrics and Statistics, JOANNA GERSTENKORN
Faculty of Economy and Sociology, University of Łódź

REFERENCES

V. CASTELLANO (1965), *Corrado Gini: a memoir*, "Metron" 24 (1-4), pp. 3-35.

E.L. CROW (1958), *The mean deviation of the Poisson distribution*, "Biometrika", 45 (3-4), pp. 556-559.

G. DALL'AGLIO (1965), *Comportamento anntotico delle stime della differenza media e del rapporto di concentrazione*, "Metron" 24 (1-4), pp. 379-414.

J.S. FRAME (1945), *Mean deviation of the binominal distribution*, "American Mathematical Monthly", 52, pp. 377-379.

T. GERSTENKORN (1971), *The recurrence relations for the moments of the discrete probability distributions*, "Dissertationes Mathematicae", 83, pp. 1-46.

T. GERSTENKORN (1975), *Bemerkungen über die zentralen unvollständigen und absoluten Momente der Pólya-Verteilung*, "Zastosowania Matematyki – Applic. Math.", 14 (4), pp. 579-597.

T. GERSTENKORN (1977), *Jednowymiarowe rozkłady dyskretne ze zniekształceniem*, in: *"Metody Statystyczne w Sterowaniu Jakością"*. Praca zbiorowa pod red. prof. Szymona Firkowicza, Ossolineum, Wrocław, Sprawozdanie z Konferencji PAN w Jabłonnej 24-28 lutego 1975 r., pp. 195-208. (*One and multivariate discrete probability distributions*, in: *"Statistical Methods in Quality Control"*, complete edition by prof. Szymon Firkowicz - Reports of the conference of the Polish Academy of Sciences at Jabłonna, November 23-28, 1975), Ossolineum, Wrocław, Polish Academy of Sciences, pp. 163-193, in Polish).

T. GERSTENKORN (1997), *The Gini's mean difference of an inflated discrete distribution*, "Proceedings of 16th Intern. Conf. on Multivariate Statistical Analysis MSA'97", November 27-29 1997, Łódź University, Chair of Statistical Methods, ed. Czesław Domański, Dariusz Parys, pp. 147-151.

C. GINI (1910), *Indici di concentrazione e di dipendenza*, "Atti della III Riunione della Societa Italiana per il progresso delle scienze", Padua 1910, pp. 453-469.

C. GINI (1911), *Variabilità e mutabilità – contributo allo studio delle distribuzioni e delle relazioni statistiche*, "Studi Economico – Giuridici della R. Università di Cagliari", vol. III, Parte II, pp. 3-159.

C. GINI (1914), *Sulla misura della concentrazione e della variabilita dei caratteri*, "Atti del R. Istituto Veneto di SS.LL.AA., a.a. 1913-1914", 73, parte II, pp. 1203-1248.

C. GINI (1921), *Measurement of inequality of incomes*, "Economic Journal" 31 (121), pp. 124-126.

C. GINI (1926), *The contribution of Italy to modern statistical methods*, "Journal of the Royal Statistical Society", 89 (4), pp. 703-724.

G.M. GIORGI (1990), *Bibliographic portrait of the Gini concentration ratio*, "Metron", 48, pp. 183-221.

G. GIRONE (1968a), *Sui momenti e sulla distribuzione della differenza media di un campione casuale di variabili esponenziali*, "Annali della Facoltà di Economia e Commercio dell'Università degli Studi di Bari", 22, pp. 97-113.

G. GIRONE (1968b), *La distribuzione della differenza media di un campione bernoulliano di variabili esponenziali*, op. ct. pp. 115-131.

N.L. JOHNSON (1957), *A note on the mean deviation of the binomial distribution*, "Biometrika", 44 (3-4), pp. 532-533.

A.R. KAMAT (1953), *The third moment of Gini's mean difference*, "Biometrika", 40 (3-4), pp. 451-452.

S.K. KATTI (1960), *Moments of the absolute difference and absolute deviation of discrete distributions*, "Annals of Mathematical Statistics", 31 (1), pp. 78-85.

M.G. KENDALL, A. STUART (1963), *The advanced theory of statistics*, Vol. 1, *Distribution Theory,* II ed., Charles Griffin & Comp., London, Sec. 2.20–2.23; Russian edition: *Теория Распределений* , Изд. "Наука", Москва 1966.

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

M.O. LORENZ (1905), *Methods of measuring the concentration of wealth*, "Journal of the American Statistical Association", *9*, p. 209.

B. MICHETTI, G. DALL'AGLIO (1957), *La differenza semplice media*, Statistica, 17 (2), pp. 159-255.

U.S. NAIR (1936), *Standard error of Gini's mean difference*, "Biometrika", 28, p. 428.

T.A. RAMASUBBAN (1956), *A χ-approximation to Gini's mean difference*, "Journal of the Indian Society of Agriculture Statistics", 8, p. 116.

T.A RAMASUBBAN (1958), *The mean difference and the mean deviation of some discontinuous distributions*, "Biometrika", 45 (3-4), pp. 549-556.

T.A. RAMASUBBAN (1959), *The generalized mean differences of the binomial and Poisson distributions*, "Biometrika", 46 (1-2), pp. 223-229.

C.R. RAO (1982), *Diversity: its measurement, decomposition, apportionment and analysis*, "Sankhyā: The Indian Journal of Statistics", Series A, 44, Part 1, pp. 1-22.

H. RINNE (1974), *Statistik I: Vorlesungsunterlagen für das Grundstudium der Wirtschaftswissenschaften im Fach Mathematik,* Band I, Justus-Liebig-Universität, Giessen,

R. RISSER, C.E. TRAYNARD (1933, 1957) *Les Principes de la Statistique Mathématique, Livre I,* Séries Statistiques, Gauthier-Villars, Paris 1933; 2ed. 1957.

G. SALVEMINI (1956), *Varianza della differenza media dei campioni ottenuti secondo lo schema di estrazione in blocco*, "Metron", 18 (1-2), pp. 133-161.
G. SALVEMINI (1957), *Distribuzione della differenza media dei campioni ricavati da una massa discreta equidistribuita*, "Atti della XVII Riunione Scientifica della Società Italiana di Statistica", Roma, pp. 69-88.
S.N. SINGH (1963), *A note on inflated Poisson distribution*, "Journal of the Indian Statistical Association", 1 (3), pp. 140-144.
M.P. SINGH (1965/66), *Inflated binomial distribution*, "Journal of Scientific Researches Banares Hindu University" 16 (1), pp. 87-90.
M.P. SINGH (1966), *A note on generalized inflated binomial distribution*, "Sankhyã: The Indian Journal of Statistics", 28 (1) p. 99.
G. UDNY YULE, M.G. KENDALL (1958), *An Introduction to the Theory of Statistics,* Charles Griffin & Co, London, 3rd ed.
G. ZANARDI (1973), *La differenza semplice media nel campione: schema con ripetizione*, "Laboratorio di Statistica - Facoltà di Economia e Commercio, Università di Venezia", pp. 1-112.
G. ZANARDI (1974), *La stima della varianza della differenza media campionaria: schema con ripetizione*, "Rivista Italiana di Economia - Demografia e Statistica", 28 (4), pp. 67-87.
R. ZIELIŃSKI (1972), *Statistical Tables*, PWN, Warsaw.

RIASSUNTO

La differenza media di Gini: teoria e applicazione alle "inflated distributions"

In questo lavoro vengono discusse alcune interessanti proprietà delle differenze medie di Gini. Il lavoro, nel quale vengono considerati sia articoli su riviste che monografie, costituisce un'importante integrazione dell'ampia rassegna, effettuata da G.M. Giorgi nel 1990, dei lavori basati sulle idee di Gini. Viene anche presentata un'applicazione delle differenze medie alle cosiddette "inflated distributions", ampiamente utilizzate nella statistica matematica.

SUMMARY

Gini's mean difference in the theory and application to inflated distributions

In the paper we give interesting properties of Gini's mean difference. We thoroughly consider the appropriate literature taking account of book publications and articles. It constitutes an important complement to the extensive bibliography of papers based on Gini's ideas, presented by G.M. Giorgi in 1990. We show an application of the mean difference to inflated distributions which are of weight and interest in statistical problems.