A χ-APPROXIMATION TO GINI'S MEAN DIFFERENCE

By T. A. Ramasubban
Indian Council of Agricultural Research, New Delhi

Introduction

One of the measures of dispersion is the Mean Difference introduced by Prof. Corrado Gini. It is the mean of all possible absolute differences of the observed values. Thus if $x_1, x_2, \cdots, x_n$ are the n observations, Gini's Mean Difference (g) is defined as

$$g = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} | x_i - x_j |. \tag{1}$$

Various authors have derived expressions for the variance of g. Nair[1] has derived an expression for its variance for samples from any population. Later, Lomnicki[2] has obtained a simpler expression for the same. Recently, Kamat[3] has evaluated the first three moments of g when the parent distribution is normal by using the absolute moments of normal distribution derived by Nabeya[4,5] and Kamat.[6] Further, on the basis of these three moments, he concludes that the statistic g may be distributed in the same manner as that of χ.

The present investigation has been carried out with a view to confirming the above conclusion arrived at by Kamat. This has been achieved by calculating the fourth moment of g and comparing its β_2 coefficients with those of χ.

2. Previous Results

Let $x_1, x_2, \cdots, x_n$ be a sample of size n from a normal population with mean μ and variance σ^2. Let $z_{ij} = x_i - x_j$ so that z_{ij} is normally distributed with mean zero and variance $2\sigma^2$, the correlation coefficient between z_{ij} and z_{kl} being defined by

$$\begin{aligned} \rho(z_{ij}, z_{kl}) &= \tfrac{1}{2} \text{ if } i = k \text{ or } j = l \\ &= -\tfrac{1}{2} \text{ if } i = l \text{ or } j = k \\ &= 0 \text{ if } i, j, k, l \text{ are all different.} \end{aligned}$$

Then, with these notations, Kamat has shown that the first three moments of g are:

$$\left.\begin{aligned} \mu_1' &= \frac{2\sigma}{\sqrt{\pi}} = 1\cdot 128379\,\sigma \\ \mu_2 &= \frac{\sigma^2}{n(n-1)}\{0\cdot 651006n + 0\cdot 151508\} \\ \text{and} \\ \mu_3 &= \frac{\sigma^3}{n^2(n-1)^2}\{0\cdot 393063n^2 + 0\cdot 399735n + 0\cdot 094822\} \end{aligned}\right\} \quad (2)$$

These moments have been evaluated by using the formulæ for absolute moments of univariate, bivariate and trivariate normal distributions.

In extending Kamat's work, we have obtained the β_2 coefficients with the help of the fourth moment, the derivation of which is dealt with in the next section.

3. Evaluation of the Fourth Moment

In order to obtain the fourth moment of g, the expectation of g^4 is necessary.

Since $z_{ij} = (x_i - x_j)$,

$$g = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} |z_{ij}| \text{ from (1).}$$

$$\begin{aligned} \therefore \quad E(g^4) = \frac{16\sigma^4}{n^4(n-1)^4} E\,[&\Sigma z_{ij}^4 + 4\{\Sigma |z_{ij}|^3 |z_{jk}| + \Sigma |z_{ij}|^3 |z_{kl}|\} \\ &+ 12\{\Sigma z_{ij}^2 |z_{jk}|\,|z_{kl}| + \Sigma |z_{ij}|\,|z_{jk}^2|\,|z_{kl}| \\ &+ \Sigma |z_{ij}|^2 |z_{kl}|\,|z_{lm}| + \Sigma |z_{ij}|\,|z_{kl}^2|\,|z_{lm}| \\ &+ \Sigma z_{ij}^2 |z_{kl}|\,|z_{mn}| + \Sigma z_{ij}^2 |z_{ik}|\,|z_{il}| \\ &+ \Sigma z_{ij}^2 |z_{ik}|\,|z_{jk}|\} + 6\{\Sigma z_{ij}^2 z_{jk}^2 + \Sigma z_{ij}^2 z_{kl}^2\} \\ &+ 24\{\Sigma |z_{ij}|\,|z_{ik}|\,|z_{jl}|\,|z_{kl}| + \Sigma |z_{ij}|\,|z_{ik}|\,|z_{jk}|\,|z_{il}| \\ &+ \Sigma |z_{ij}|\,|z_{ik}|\,|z_{jk}|\,|z_{lm}| + \Sigma |z_{ij}|\,|z_{ik}|\,|z_{jl}|\,|z_{im}| \\ &+ \Sigma |z_{ij}|\,|z_{ik}|\,|z_{il}|\,|z_{im}| + \Sigma |z_{ij}|\,|z_{jk}|\,|z_{kl}|\,|z_{lm}| \\ &+ \Sigma |z_{ij}|\,|z_{ik}|\,|z_{jl}|\,|z_{mn}| + \Sigma |z_{ij}|\,|z_{ik}|\,|z_{il}|\,|z_{mn}| \\ &+ \Sigma |z_{ij}|\,|z_{jk}|\,|z_{lm}|\,|z_{mn}| + \Sigma |z_{ij}|\,|z_{ik}|\,|z_{lm}|\,|z_{np}| \\ &+ \Sigma |z_{ij}|\,|z_{kl}|\,|z_{mn}|\,|z_{pq}|\}] \end{aligned} \quad (3)$$

where i, j, k, l, m, n, p and q are all different and the summation is taken over all of them with the restriction that the first subscript of z is always less than the second subscript.

The different types of terms occurring in (3) and the number of times each term occurs are given in Table I.

TABLE I

Sl. No.	Type of Term	Number of Terms
1	z_{ij}^4	$\frac{1}{2}\, n^{[2]}$
2	$\lvert z_{ij}\rvert^3 \lvert z_{jk}\rvert$	$n^{[3]}$
3	$\lvert z_{ij}\rvert^3 \lvert z_{kl}\rvert$	$\frac{1}{4}\, n^{[4]}$
4	$z_{ij}^2 \lvert z_{jk}\rvert \lvert z_{kl}\rvert$	$n^{[4]}$
5	$\lvert z_{ij}\rvert z_{jk}^2 \lvert z_{kl}\rvert$	$\frac{1}{2}\, n^{[4]}$
6	$z_{ij}^2 \lvert z_{kl}\rvert \lvert z_{lm}\rvert$	$\frac{1}{4}\, n^{[5]}$
7	$\lvert z_{ij}\rvert z_{kl}^2 \lvert z_{lm}\rvert$	$\frac{1}{2}\, n^{[5]}$
8	$z_{ij}^2 \lvert z_{kl}\rvert \lvert z_{mn}\rvert$	$\frac{1}{16}\, n^{[6]}$
9	$z_{ij}^2 \lvert z_{ik}\rvert \lvert z_{il}\rvert$	$\frac{1}{2}\, n^{[4]}$
10	$z_{ij}^2 \lvert z_{ik}\rvert \lvert z_{jk}\rvert$	$\frac{1}{2}\, n^{[3]}$
11	$z_{ij}^2 z_{jk}^2$	$\frac{1}{2}\, n^{[3]}$
12	$z_{ij}^2 z_{kl}^2$	$\frac{1}{8}\, n^{[4]}$
13	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{jl}\rvert \lvert z_{kl}\rvert$	$\frac{1}{8}\, n^{[4]}$
14	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{jk}\rvert \lvert z_{il}\rvert$	$\frac{1}{2}\, n^{[4]}$
15	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{jk}\rvert \lvert z_{lm}\rvert$	$\frac{1}{12}\, n^{[5]}$
16	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{jl}\rvert \lvert z_{im}\rvert$	$\frac{1}{2}\, n^{[5]}$
17	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{il}\rvert \lvert z_{im}\rvert$	$\frac{1}{24}\, n^{[5]}$
18	$\lvert z_{ij}\rvert \lvert z_{jk}\rvert \lvert z_{kl}\rvert \lvert z_{lm}\rvert$	$\frac{1}{2}\, n^{[5]}$
19	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{jl}\rvert \lvert z_{mn}\rvert$	$\frac{1}{4}\, n^{[6]}$
20	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{il}\rvert \lvert z_{mn}\rvert$	$\frac{1}{12}\, n^{[6]}$
21	$\lvert z_{ij}\rvert \lvert z_{jk}\rvert \lvert z_{lm}\rvert \lvert z_{mn}\rvert$	$\frac{1}{8}\, n^{[6]}$
22	$\lvert z_{ij}\rvert \lvert z_{ik}\rvert \lvert z_{lm}\rvert \lvert z_{np}\rvert$	$\frac{1}{16}\, n^{[7]}$
23	$\lvert z_{ij}\rvert \lvert z_{kl}\rvert \lvert z_{mn}\rvert \lvert z_{pq}\rvert$	$\frac{1}{384}\, n^{[8]}$

Note.—$n^{[i]} = n(n-1)(n-2)\ldots\ldots\ldots(n-\overline{i-1})$.

The expected values of the first twelve types of summations can be calculated from the following formulæ developed by Kamat:—

$$\left.\begin{aligned}
(4, 0, 0) &= 3 \\
(3, 1, 0) &= \frac{2}{\pi}\{\sqrt{(1-\rho_{12}^2)}\,(2+\rho_{12}^2) + 3\rho_{12}\sin^{-1}\rho_{12}\} \\
(2, 2, 0) &= 1 + 2\rho_{12}^2 \\
(2, 1, 1) &= \frac{2}{\pi}\{(\rho_{23} + 2\rho_{12}\rho_{13})\sin^{-1}\rho_{23} \\
&\qquad + \sqrt{(1-\rho_{23}^2)}\,(1+\rho_{12}^2+\rho_{13}^2)\}
\end{aligned}\right\} \quad (4)$$

where,

$$(l_1, l_2, l_3, \cdots, l_k) = \int \int \int_{-\infty}^{\infty} \cdots \int | y_1 |^{l_1} | y_2 |^{l_2} | y_3 |^{l_3} \cdots | y_k |^{l_k}$$
$$\times p(y_1, y_2, \cdots, y_k)\, dy_1\, dy_2 \cdots dy_k$$

$p(y_1, y_2, \cdots, y_k)$ denoting the k-variate normal distribution and ρ_{ij} is the correlation coefficient between y_i and y_j.

To evaluate the expectations of the last eleven types of sums, the series for (1, 1, 1, 1) given by Kamat upto the fourth power of correlation-coefficients has been extended by us upto the ninth power of ρ^8.

In its extended form, the above series becomes

$$\frac{4}{\pi^2}\Big\{1 + \tfrac{1}{2}\sum \rho_{ij}^2 + \sum \rho_{ij}\rho_{ik}\rho_{jk} + \frac{1}{24}\sum \rho_{ij}^4 - \tfrac{1}{4}\sum \rho_{ij}^2\rho_{ik}^2$$
$$+ \tfrac{1}{4}\sum \rho_{ij}^2\rho_{kl}^2 + \sum \rho_{ij}\rho_{ik}\rho_{kl}\rho_{jl} + \tfrac{1}{6}\sum \rho_{ij}^3\rho_{ik}\rho_{jk}$$
$$+ \frac{1}{80}\sum \rho_{ij}^6 - \frac{1}{16}\sum \rho_{ij}^4\rho_{ik}^2 + \frac{1}{48}\sum \rho_{ij}^4\rho_{kl}^2 + \tfrac{1}{8}\sum \rho_{ij}^2\rho_{jk}^2\rho_{kl}^2$$
$$+ \tfrac{3}{8}\sum \rho_{ij}^2\rho_{ik}^2\rho_{il}^2 - \tfrac{1}{8}\sum \rho_{ij}^2\rho_{ik}^2\rho_{jk}^2 + \tfrac{1}{6}\sum \rho_{ij}^3\rho_{ik}\rho_{jl}\rho_{kl}$$
$$+ \tfrac{1}{2}\sum \rho_{ij}^2\rho_{ik}\rho_{il}\rho_{jk}\rho_{jl} + \frac{3}{40}\sum \rho_{ij}^5\rho_{ik}\rho_{jk} + \frac{1}{12}\sum \rho_{ij}^3\rho_{ik}^3\rho_{jk}$$
$$- \tfrac{1}{4}\sum \rho_{ij}^2\rho_{ik}^2\rho_{jk}\rho_{jl}\rho_{kl} + \cdots\Big\} \tag{5}$$

On substituting the corresponding expectations calculated from (4) and (5) in expression (3) and remembering the number of terms shown in the last column of Table I, we find that

$$E(g^4) = \frac{64\sigma^4}{n^3(n-1)^3\pi^2}\Big\{\frac{n^6}{4} + 0\cdot01694n^5 - 0\cdot10195n^4$$
$$- 1\cdot27089n^3 + 15\cdot78622n^2 - 55\cdot88448n$$
$$+ 58\cdot68473\Big\} \tag{6}$$

whence,

$$\mu_4 = \frac{64\sigma^4}{n^3(n-1)^3\pi^2}\{0\cdot22986n^4 - 1\cdot43549n^3 + 15\cdot81994n^2$$
$$- 55\cdot81847n + 58\cdot68473\}$$

$$= \frac{\sigma^4}{n^3 (n-1)^3} \{1 \cdot 49054 n^4 - 9 \cdot 30851 n^3 + 102 \cdot 58528 n^2 - 361 \cdot 95798 n + 380 \cdot 54440\} \quad (7)$$

Using the moments obtained in (2) and (7), the β_2 coefficients for g have been determined for $n = 5$, 10, 15 and 20. The corresponding values of β_2 have also been found for the χ-distribution having the same coefficient of variation as that of g. These values are shown in Table II.

TABLE II

Sample size (n)	For g		For χ-distribution	
	β_1*	β_2	β_1*	β_2
5	0·1797	3·892	0·1672	3·061
10	0·0708	3·159	0·0648	3·010
15	0·0436	3·006	0·0398	3·005
20	0·0315	3·006	0·0286	3·005

* Values obtained by Kamat.

4. Discussion of the Results

Allowing for the error of approximating the series (5), the β_2 values of g shown in Table II seem to be in very close agreement with those of χ for samples of size greater than ten. For $n = 5$ and 10, the differences observed in β_2 for g and χ are not statistically significant. Also, β_1 values of g based upon exact formulæ and those of χ are nearly the same. Thus, the above results lead to the conclusion that at least for $n > 10$, g follows a χ-distribution having the degrees of freedom (ν) given by the relation:

$$\frac{\nu \left(\left| \frac{\nu}{2} \right. \right)^2}{\left(\left| \frac{\nu+1}{2} \right. \right)^2} = \frac{2}{n(n-1)} \{n^2 - 0 \cdot 488701 n + 0 \cdot 118994\} \quad (8)$$

The distribution of g for samples of size less than ten is being investigated and will be reported later.

5. Summary

Following the method used by Kamat to obtain the first three moments of Gini's Mean Difference (g), we have evaluated its fourth moment and obtained therefrom the β_2 coefficients. On comparison of these β_2 values with those of a χ-distribution having the same coefficient of variation, it is possible to confirm the conjecture made by Kamat that the distribution of g may be quite close to that of χ.

6. Acknowledgment

Before concluding, I would like to express my indebtedness to Dr. B. V. Sukhatme, Indian Council of Agricultural Research, for his very valuable advice and suggestions during the preparation of this paper.

References

1. Nair, U. S. .. *Biometrika*, 1936, **28**, 428.
2. Lomnicki, Z. A. .. *Ann. Math. Stat.*, 1952, **23**, 635.
3. Kamat, A. R. .. *Biometrika*, 1953, **40**, 451.
4. Nabeya, S. .. *Ann. Inst. Stat. Math.*, 1951, **3**, 2.
5. ——— .. *Ibid.*, 1952, **4**, 15.
6. Kamat, A. R. .. *Biometrika*, 1953, **40**, 20.