ACTA MATHEMATICA VIETNAMICA
Volume 36, Number 3, 2011, pp. 685–694


# ON THE GINI MEAN DIFFERENCE

PHAM XUAN BINH

ABSTRACT. Gini mean difference (GMD) was proposed by Gini in 1912 for measuring income inequalities or economic wealth. The aim of this paper is to prove an estimation of this GMD in the general case of distribution.

## 1. Introduction

Many achievements have been made since the first article on Gini mean difference (GMD) by Italian mathematician Corrado Gini (1912). During the last more than 80 years, GMD became one of the most popular inequalities measure used in economics. The mean difference is a measure of statistical dispersion which equals the average absolute difference of two independent random values $X$ and $X'$ drawn from a probability distribution $F$: $E|X - X'|$.

Up to now, many results on bounds for $E|X - X'|$ have been obtained in many special cases of the distribution $F$ (see, for example, [1]-[5]). In this paper, we attempt to give a proof of the following inequality in the general case of the distribution $F$:

$$E|X - X'| \leq \frac{2}{\sqrt{3}}\sqrt{Var(X)}.$$

Also, some identities related to GMD are investigated.

## 2. Main results

Let $X$ be a random variable with the distribution function $F$ such that $EX^2 < +\infty$. Let us denote by $X'$ a random variable that is independent of $X$ and has the same distribution as $X$. For the distribution function $F(x) = P(X \leq x)$, we define the function $G$ by

$$G(x) = \frac{F(x) + F(x^-)}{2},\ x \in \mathbb{R}.$$

Received May 25, 2010; in revised form September 7, 2010.
2000 *Mathematics Subject Classification.* 60E15, 62H20.
*Key words and phrases.* Cauchy-Schwarz inequality, distribution function, Gini mean difference, Hermite-Hadamard inequality, Lebesgue decomposition theorem.

**Lemma 2.1** (Hermite-Hadamard inequality, see [8], [9]). *Let $f : [a, b) \to \mathbb{R}$ be a convex function. Then,*

$$f\left(\frac{a+b}{2}\right) \le \frac{1}{b-a}\int_a^b f(x)dx \left(\le \frac{f(a)+f(b)}{2}\right).$$

**Lemma 2.2.** *Let $F = F_d$ be a pure jump distribution function. Then:*

(i) $EG(X) = \dfrac{1}{2}$,

(ii) *for $q > 1$ we have $E[G(X)]^q \le \dfrac{1}{q+1}$.*

*Proof.* (i) We know that $F = F_d$ has at most a countable number of jumps. Let $\{x_k : k \in \mathbb{N}\}$ be those jumps. Then,

$$\begin{aligned} EG(X) &= \frac{1}{2}\int_{-\infty}^{+\infty} [F(x) + F(x^-)]dF(x) \\ &= \frac{1}{2}\sum_k [F(x_k) + F(x_k^-)][F(x_k) - F(x_k^-)] \\ &= \frac{1}{2}\sum_k [F^2(x_k) - F^2(x_k^-)] \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} d\left[F^2(x)\right] = \frac{1}{2}. \end{aligned}$$

(ii) By Lemma 2.1, we have

$$\begin{aligned} E[G(X)]^q &= \frac{1}{2}\int_{-\infty}^{+\infty} \left[\frac{F(x) + F(x^-)}{2}\right]^q dF(x) \\ &= \sum_k \left[\frac{F(x_k) + F(x_k^-)}{2}\right]^q [F(x_k) - F(x_k^-)] \\ &\le \sum_k \int_{F(x_k^-)}^{F(x_k)} y^q dy = \int_0^1 y^q dy = \frac{1}{q+1}. \end{aligned}$$

□

**Lemma 2.3.** *Let $F = F_c$ be a continuous distribution function. Then, for $q \ge 1$ we have*

$$E[G(X)]^q = \frac{1}{q+1}.$$

*Proof.* The formula for partial integration (see [7], page 66) gives us

$$E[G(X)]^q = \int_{-\infty}^{+\infty} [F(x)]^q dF(x)$$

$$= [F(x)]^{q+1}\Big|_{-\infty}^{+\infty} - q\int_{-\infty}^{+\infty} [F(x)]^q dF(x)$$

$$= 1 - qE[G(X)]^q.$$

□

Now we return to the general case.

**Proposition 2.4.** *The following statements hold:*

(i) $EG(X) = \dfrac{1}{2}$;

(ii) $E[G(X)]^2 \leq \dfrac{1}{3}$;

(iii) $Var\{G(X)\} \leq \dfrac{1}{12}$.

*Proof.* By the Lebesgue Decomposition Theorem we see that $F$ can be uniquely written as a convex combination of a discrete distribution function $F_d$ and a continuous distribution function $F_c$, i.e., $F = \alpha F_d + (1-\alpha)F_c$ where $0 \leq \alpha \leq 1$.
(i) At first, by continuity of $F_c$ and the formula for partial integration, we obtain

$$\int_{-\infty}^{+\infty} \frac{F_d(x)+F_d(x^-)}{2} dF_c(x) + \int_{-\infty}^{+\infty} F_c(x)dF_d(x)$$

$$= \int_{-\infty}^{+\infty} F_d(x)dF_c(x) + \int_{-\infty}^{+\infty} F_c(x)dF_d(x)$$

$$= \int_{-\infty}^{+\infty} d[F_c(x)F_d(x)] = 1.$$

Now, by Lemma 2.1, Lemma 2.2, and

$$G(x) = \alpha\frac{F_d(x)+F_d(x^-)}{2} + (1-\alpha)F_c(x),$$

we have

$$EG(X) = \int_{-\infty}^{+\infty} \frac{F(x)+F(x^-)}{2} dF(x)$$

$$= \int_{-\infty}^{+\infty} \left[\alpha\frac{F_d(x)+F_d(x^-)}{2} + (1-\alpha)F_c(x)\right] d[\alpha F_d(x) + (1-\alpha)F_c(x)]$$

$$= \alpha^2 \int_{-\infty}^{+\infty} \frac{F_d(x) + F_d(x^-)}{2} dF_d(x) + (1-\alpha)^2 \int_{-\infty}^{+\infty} F_c(x) dF_c(x)$$

$$+ \alpha(1-\alpha) \int_{-\infty}^{+\infty} \frac{F_d(x) + F_d(x^-)}{2} dF_c(x) + (1-\alpha)\alpha \int_{-\infty}^{+\infty} F_c(x) dF_d(x)$$

$$= \frac{\alpha^2}{2} + \frac{(1-\alpha)^2}{2} + \alpha(1-\alpha) \int_{-\infty}^{+\infty} F_d(x) dF_c(x) + \alpha(1-\alpha) \int_{-\infty}^{+\infty} F_c(x) dF_d(x)$$

$$= \frac{\alpha^2}{2} + \frac{(1-\alpha)^2}{2} + \alpha(1-\alpha) = \frac{1}{2}.$$

(ii) By Lemma 2.1, Lemma 2.2 and the formula for partial integration we have

$$E[G(X)]^2 = \int_{-\infty}^{+\infty} \left[\alpha \frac{F_d(x) + F_d(x^-)}{2} + (1-\alpha) F_c(x)\right]^2 d[\alpha F_d(x) + (1-\alpha) F_c(x)]$$

$$= \int_{-\infty}^{+\infty} \Big\{\alpha^2 \left[\frac{F_d(x) + F_d(x^-)}{2}\right]^2 + (1-\alpha)^2 F_c^2(x)$$

$$+ 2\alpha(1-\alpha) \left[\frac{F_d(x) + F_d(x^-)}{2}\right] F_c(x)\Big\} d[\alpha F_d(x) + (1-\alpha) F_c(x)]$$

$$= \alpha^3 \int_{-\infty}^{+\infty} \left[\frac{F_d(x) + F_d(x^-)}{2}\right]^2 dF_d(x)$$

$$+ \alpha^2(1-\alpha) \int_{-\infty}^{+\infty} \left[\frac{F_d(x) + F_d(x^-)}{2}\right]^2 dF_c(x)$$

$$+ (1-\alpha)^2 \alpha \int_{-\infty}^{+\infty} F_c^2(x) dF_d(x) + (1-\alpha)^3 \int_{-\infty}^{+\infty} F_c^2(x) dF_c(x)$$

$$+ 2\alpha^2(1-\alpha) \int_{-\infty}^{+\infty} \left[\frac{F_d(x) + F_d(x^-)}{2}\right] F_c(x) dF_d(x)$$

$$+ 2\alpha(1-\alpha)^2 \int_{-\infty}^{+\infty} \left[\frac{F_d(x) + F_d(x^-)}{2}\right] F_c(x) dF_c(x)$$

$$= \alpha^3 \int_{-\infty}^{+\infty} \left[\frac{F_d(x) + F_d(x^-)}{2}\right]^2 dF_d(x) + \alpha^2(1-\alpha) \int_{-\infty}^{+\infty} F_d^2(x) dF_c(x)$$

$$+ (1-\alpha)^2\alpha \int\limits_{-\infty}^{+\infty} F_c^2(x)dF_d(x) + \frac{(1-\alpha)^3}{3}$$

$$+ 2\alpha^2(1-\alpha) \int\limits_{-\infty}^{+\infty} \left[\frac{F_d(x)+F_d(x^-)}{2}\right] F_c(x)dF_d(x)$$

$$+ 2\alpha(1-\alpha)^2 \int\limits_{-\infty}^{+\infty} F_d(x)F_c(x)dF_c(x). \tag{2.1}$$

We now give the upper bound and compute the remaining five terms on the right hand side of (2.1):

$$\alpha^3 \int\limits_{-\infty}^{+\infty} \left[\frac{F_d(x)+F_d(x^-)}{2}\right]^2 dF_d(x) \le \frac{\alpha^3}{3}, \tag{2.2}$$

$$\alpha^2(1-\alpha) \int\limits_{-\infty}^{+\infty} F_d^2(x)dF_c(x) = \alpha^2(1-\alpha) - \alpha^2(1-\alpha) \int\limits_{-\infty}^{+\infty} F_c(x)dF_d^2(x)$$

$$= \alpha^2(1-\alpha) - \alpha^2(1-\alpha)\sum_k F_c(x_k)\left[F_d^2(x_k) - F_d^2(x_k^-)\right], \tag{2.3}$$

$$2\alpha^2(1-\alpha) \int\limits_{-\infty}^{+\infty} \left[\frac{F_d(x)+F_d(x^-)}{2}\right] F_c(x)dF_d(x) =$$

$$= 2\alpha^2(1-\alpha)\sum_k \frac{F_d(x_k)+F_d(x_k^-)}{2} F_c(x_k)[F_d(x_k) - F_d(x_k^-)]$$

$$= \alpha^2(1-\alpha)\sum_k F_c(x_k)[F_d^2(x_k) - F_d^2(x_k^-)], \tag{2.4}$$

$$(1-\alpha)^2\alpha \int\limits_{-\infty}^{+\infty} F_c^2(x)dF_d(x) = \alpha(1-\alpha)^2 \sum_k F_c^2(x_k)[F_d(x_k) - F_d(x_k^-)], \tag{2.5}$$

$$2\alpha(1-\alpha)^2 \int\limits_{-\infty}^{+\infty} F_d(x)F_c(x)dF_c(x) = \alpha(1-\alpha)^2 \int\limits_{-\infty}^{+\infty} F_d(x)\left[dF_c^2(x)\right]$$

$$= \alpha(1-\alpha)^2 - \alpha(1-\alpha)^2 \int\limits_{-\infty}^{+\infty} F_c^2(x)dF_d(x)$$

$$= \alpha(1-\alpha)^2 - \alpha(1-\alpha)^2 \sum_k F_c^2(x_k)[F_d(x_k) - F_d(x_k^-)]. \tag{2.6}$$

From (2.1), (2.2), (2.3), (2.4), (2.5) and (2.6) we obtain

$$\begin{aligned} E[G(X)]^2 &\leq \frac{\alpha^3}{3} + \alpha^2(1-\alpha) + \alpha(1-\alpha)^2 + \frac{(1-\alpha)^3}{3} \\ &\leq \frac{\alpha^3}{3} + \alpha^2 - \alpha^3 + \alpha + \alpha^3 - 2\alpha^2 + \frac{1}{3} - \frac{3\alpha}{3} + \frac{3\alpha^2}{3} - \frac{\alpha^3}{3} \\ &\leq \frac{1}{3}. \end{aligned}$$

(iii) $VarG(X) = E[G(X)]^2 - [EG(X)]^2 \leq \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$ □

**Proposition 2.5.** *Let $g$ be a non-decreasing function on $\mathbb{R}$ such that $Eg(X) < +\infty$. Then,*

$$E|g(X) - g(X')| = 4E\left\{ g(X)\left[ G(X) - \frac{1}{2} \right] \right\},$$

*where $G(x) = \dfrac{F(x) + F(x^-)}{2}$ for all $x \in \mathbb{R}$.*

*Proof.* We have

$$\begin{aligned} E|g(X) - g(X')| &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |g(u) - g(v)| dF(u)dF(v) \\ &= \iint_{u<v} [g(v) - g(u)] dF(u)dF(v) + \iint_{u>v} [g(u) - g(v)] dF(u)dF(v) \\ &= 2\iint_{u<v} [g(v) - g(u)] dF(u)dF(v) \\ &= 2\iint_{u<v} g(v) dF(u)dF(v) - 2\iint_{u<v} g(u) dF(u)dF(v) \\ &= 2\int_{-\infty}^{+\infty} g(v) \int_{-\infty}^{v^-} dF(u)dF(v) - 2\int_{-\infty}^{+\infty} g(u) \int_{u^+}^{+\infty} dF(v)dF(u) \\ &= 2\int_{-\infty}^{+\infty} g(v)F(v^-)dF(v) - 2\int_{-\infty}^{+\infty} g(u)\left[1 - F(u^+)\right] dF(u) \\ &= 2\int_{-\infty}^{+\infty} g(x)F(x^-)dF(x) - 2\int_{-\infty}^{+\infty} g(x)\left[1 - F(x^+)\right] dF(x) \end{aligned}$$

$$= 2 \int_{-\infty}^{+\infty} g(x) \left[F(x^+) + F(x^-) - 1\right] dF(x)$$
$$= 4 \int_{-\infty}^{+\infty} g(x) \left[\frac{F(x^+) + F(x^-)}{2} - \frac{1}{2}\right] dF(x)$$
$$= 4 \int_{-\infty}^{+\infty} g(x) \left[\frac{F(x) + F(x^-)}{2} - \frac{1}{2}\right] dF(x)$$
$$= 4 \int_{-\infty}^{+\infty} g(x) \left[G(x) - \frac{1}{2}\right] dF(x) = 4E\left\{g(X)\left[G(X) - \frac{1}{2}\right]\right\}.$$

□

**Corollary.** *Under the assumption stated in Proposition 2.5, the following equalities hold:*

(i)

$$\begin{aligned} E|g(X) - g(X')| &= 4E[g(X) - Eg(X)]G(X) \\ &= 4E[g(X) - Eg(X)][G(X) - a] \\ &= 4E[g(X) - b][G(X) - \frac{1}{2}] \end{aligned}$$

*for all* $a, b \in \mathbb{R}$.

(ii)

$$\begin{aligned} E|X - X'| &= 4E\left(X[G(X) - \frac{1}{2}]\right) \\ &= 4E\left[(X - EX)\, G(X)\right] \\ &= 4E\left((X - EX)\left[G(X) - \frac{1}{2}\right]\right) \\ &= 4E[X - EX][G(X) - a] \\ &= 4E[X - b][G(X) - \frac{1}{2}] \end{aligned}$$

*for all* $a, b \in \mathbb{R}$.

We are now in a position to state the main result.

**Theorem 2.6.**

$$E|X - X'| \leq \frac{2}{\sqrt{3}}\sqrt{Var(X)}.$$

*Equality holds if and only if* $X$ *has a uniform distribution on some interval* $[a, b]$.

*Proof.* By the Cauchy-Schwarz inequality, the Corollary of Proposition 2.5 and part (iii) of Proposition 2.4, we obtain

$$\begin{aligned} E|X - X'| &= 4E[X - EX][G(X) - \frac{1}{2}] \\ &= 4E[X - EX][G(X) - EG(X)] \\ &\le 4\sqrt{VarX.VarG(X)} \\ &\le 4\sqrt{\frac{1}{12}}\sqrt{Var(X)} \\ &\le \frac{2}{\sqrt{3}}\sqrt{Var(X)}. \end{aligned}$$

It is clear that equality holds if and only if $X$ and $G(X)$ are linearly dependent, that is, if and only if $X$ has a uniform distribution on some interval $[a, b]$. □

We give now some useful identities for the Gini mean difference.

**Proposition 2.7.**

$$E|X - X'| = 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - y)[G(x) - G(y)]dF(x)dF(y).$$

*Proof.* By part (ii) of the Corollary of Proposition 2.5 we have

$$\begin{aligned} &2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - y)[G(x) - G(y)]dF(x)dF(y) \\ =&2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xG(x)dF(x)dF(y) - 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yG(x)dF(x)dF(y) \\ &- 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xG(y)dF(x)dF(y) + 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yG(y)dF(x)dF(y) \\ =&4\int_{-\infty}^{+\infty}xG(x)dF(x) - 4\int_{-\infty}^{+\infty}xdF(x)\int_{-\infty}^{+\infty}G(y)dF(y) \\ =&4EXG(X) - 2EX = 4EX[G(X) - \frac{1}{2}] = E|X - X'|. \end{aligned}$$

□

**Proposition 2.8.**

$$E|X - X'| = 2\int_{-\infty}^{+\infty}F(x)[1 - F(x)]dx = 2\int_{-\infty}^{+\infty}G(x)[1 - G(x)]dx.$$

*Proof.* By the equality $\max\{X,X'\} = \frac{1}{2}(X+X') + \frac{1}{2}|X-X'|$ and since $\max\{X,X'\}$ has the distribution function $F^2$, we have

$$\begin{aligned}
E|X-X'| &= 2E\left(\max\{X,X'\}\right) - 2EX \\
&= 2\int_{-\infty}^{+\infty} x dF^2(x) - 2\int_{-\infty}^{+\infty} x dF(x) \\
&= 2\int_{0}^{+\infty} [1-F^2(x)]dx - 2\int_{-\infty}^{0} F^2(x)dx \\
&\quad - 2\int_{0}^{+\infty} [1-F(x)]dx + 2\int_{-\infty}^{0} F(x)dx \\
&= 2\int_{0}^{+\infty} [1-F(x)]F(x)dx + 2\int_{-\infty}^{0} F(x)[1-F(x)]dx \\
&= 2\int_{-\infty}^{+\infty} F(x)[1-F(x)]dx.
\end{aligned}$$

On the other hand, since the set $\{x \in \mathbb{R} : F(x) \neq G(x)\}$ has at most a countable number of points, we have

$$\int_{-\infty}^{+\infty} F(x)[1-F(x)]dx = \int_{-\infty}^{+\infty} G(x)[1-G(x)]dx.$$

□

We conclude the paper with the following note. As pointed out in Proposition 2.8 and the Corollary of Proposition 2.5, we obtain the identity

$$\begin{aligned}
E|X-X'| &= 2\int_{-\infty}^{+\infty} F(x)[1-F(x)]dx = 4E\left(X[G(X)-\frac{1}{2}]\right) \\
&= 4\int_{-\infty}^{+\infty} x\left[G(x)-\frac{1}{2}\right]dF(x) = 2\int_{-\infty}^{+\infty} x[2G(x)-1]dF(x).
\end{aligned}$$

When the distribution function $F$ is continuous, we can use the formula for partial integration to get

$$E|X-X'| = 2\int_{-\infty}^{+\infty} F(x)[1-F(x)]dx = 2\int_{-\infty}^{+\infty} x[2F(x)-1]dF(x).$$

However, in general it frequently happens that

$$E|X - X'| = 2 \int_{-\infty}^{+\infty} F(x)[1 - F(x)]dx \neq 2 \int_{-\infty}^{+\infty} x[2F(x) - 1]dF(x),$$

because the formula for partial integration can not be applied (see [4], page 883). For example, if $P(X = 0) = P(X = 1) = \frac{1}{2}$ then it is easy to check that

$$E|X - X'| = 2 \int_{-\infty}^{+\infty} F(x)[1 - F(x)]dx = 2 \int_{-\infty}^{+\infty} G(x)[1 - G(x)]dx$$
$$= 2 \int_{-\infty}^{+\infty} x[2G(x) - 1]dF(x) = \frac{1}{2},$$

while

$$2 \int_{-\infty}^{+\infty} x[2F(x) - 1]dF(x) = 2(2 - 1)\frac{1}{2} = 1.$$

## References

[1] P. Cerone, Bounding the Gini Mean Difference, *International Series of Numerical Mathematics*, **157** (2008), 77-89.
[2] P. Cerone and S. S. Dragomir, Bounds for the Gini Mean Difference via the Sonin identity, *Comp. Math. Appl.* **50**(3/4) (2005), 599-609.
[3] P. Cerone and S. S. Dragomir, Bounds for the $r$-weighted Gini Mean Difference of an empirical distribution, *Comp. Math. Appl.* **49** (2009), 180-188.
[4] P. C. Chang and E. A. Johnson, Some distribution-free properties of the asymptotic variance of the Spearman estimator in bioassays, *Biometrics* **28** (1972), 882-889.
[5] H. A. David, Gini's mean difference rediscovered, *Biometrika* **55** (1968), 573.
[6] C. Gini, Variabilit e Mutabilit, contributo allo studio delle distribuzioni e relazioni statistiche, *Studi Economico-Giuridici dell'Univ. di Cagliari* (Part 2) (1912), 1-158.
[7] Gut A., *Probability: A Graduate Course*, Springer, 2007.
[8] D. S. Mitrinovc, J. E. Pecarc, and A. M. Fink, *Classical and New Inequalities in Analysis*, Kluwer Academic Publishers, Dodrecht/Boston/London, 1993.
[9] C. P. Nicules and L. E. Persson, *Convex Function and Their Applications - A contemprorary Approach*, Springer, 2006.

Department of Mathematics, Quy Nhon University
170 An Duong Vuong, Quy Nhon, Binh Dinh, Vietnam.
*E-mail address*: phamxuanbinh@qnu.edu.vn, lamthithanhtam@qnu.edu.vn