Sharp Inequalities for Expectations of Products of Two Random Variables

Chul Gyu Park[a] and Seokhoon Yun[b]

[a] *School of Mathematics and Statistics, Carleton University, Canada K1S 5B6*

[b] *Department of Applied Statistics, University of Suwon, Korea 445-743*

Abstract

In this article we provide lower and upper bounds for the expectation of the product of two random variables. These bounds are sharper than those in the Cauchy-Schwarz inequality and even in the Hölder inequality. We also derive a monotone property for the covariance between infinitely divisible random variables. Our results might be used in checking whether the correlation structure assumed for the simulation is consistent.

Keywords: Sharp bounds; Cauchy-Schwarz inequality; Hölder inequality; Covariances; Infinitely divisible random variables; Simulation

AMS 2000 Subject Classifications: Primary 60E15; secondary 65C05

1 Introduction

In many recent studies, more than one response variable is assumed to be observed on each of the experimental units. One example of this kind of data is found in Regan and Catalano (1999) where they analyze the joint risk of malformation and low fetal weights incurred by increased dosage levels. Another interesting example is the set of data obtained by Feagan et al. (1995) to evaluate the antiflammatory drug methotrexate for patients with chronically active Crohn's disease. Here the efficacy of the drug was measured by using two instruments: the Crohn's Disease Activity Index and the Inflammatory Bowel Disease Questionnaire.

Since there might be nonignorable dependence among the response variables if they come from the same subject, it is generally hard to construct the likelihood function based on the observed data. For analyzing such data, we usually rely on the generalized estimating equations (GEE) approach which is a variant of the quasi-likelihood method. This approach simply specifies marginal models for the outcome variables and their correlation structure instead of specifying the complete joint distribution for those dependent observations. One may refer to Liang and Zeger (1986) and Zeger, Liang and Albert (1988) for details of GEE.

Although asymptotic properties of GEE estimators are well known, their finite-sample behaviors are generally unknown and must be investigated through a Monte Carlo simulation. For this, we have to specify the underlying parameters upon which investigations will be made. However, unlike the parameters in marginal models, correlations among the outcome variables cannot be specified arbitrarily between -1 and 1 if they have different families of distributions. This can be easily understood by noting that a normal and a Poisson variable can never have a correlation coefficient of 1. Consequently, it may well be interesting to find the upper and lower bounds for the correlation coefficient between the random variables when their marginal distributions are specified as different.

Once marginal distributions are specified, our problem is equivalent to finding the bounds for the expectation of the product of two random variables. As we shall see in the next section, our results are stronger than the Caushy-Schwarz inequality and even the Hölder inequality.

2 Main Results

For two random variables X and Y with finite second moments, the Cauchy-Schwarz inequality says

$$|E(XY)| \leq [E(X^2)E(Y^2)]^{1/2},$$

where the equality holds if and only if there exist real numbers a and b not both zero such that $P\{aX + bY = 0\} = 1$. But, there might be a question about the bound when X and Y are not proportional to each other or when their marginal distributions are of different types. In this section we provide lower and upper bounds for $E(XY)$ which are sharper than the Cauchy-Schwarz bounds.

Throughout this section all the random variables referred to are assumed to be defined on a fixed probability space $(\Omega, \mathcal{F}, P)$. A random variable X is said to be simple if it can be represented as a finite sum of the form $X = \sum_i x_i I_{A_i}$, $x_i \in \mathbb{R}$, $A_i \in \mathcal{F}$, where $\{A_i\}$ forms a finite partition of Ω. I_A denotes the indicator function of a set A. For a distribution function F, let F^{-1} denote the quantile function of F. We begin with the case of simple random variables.

Lemma 2.1 *For simple random variables X and Y with distribution functions F and G, respectively, we have*

$$E(F^{-1}(U)G^{-1}(1-U)) \leq E(XY) \leq E(F^{-1}(U)G^{-1}(U)), \tag{2.1}$$

where U is a Uniform(0,1) random variable.

Proof. Write $X = \sum_{i=1}^m x_i I_{A_i}$, $A_i = \{X = x_i\}$ and $Y = \sum_{j=1}^n y_j I_{B_j}$, $B_j = \{Y = y_j\}$, where $\{A_i\}$ and $\{B_j\}$ form finite partitions of Ω. We may assume $-\infty < x_1 < x_2 < \cdots < x_m < \infty$ and $-\infty < y_1 < y_2 < \cdots < y_n < \infty$. Also, we write

$$\begin{aligned} p_{ij} &= P\{X = x_i, Y = y_j\},\ i = 1, ..., m;\ j = 1, ..., n, \\ q_i &= P\{X \leq x_i\} = \sum_{k=1}^{i} \sum_{j=1}^{n} p_{kj},\ i = 1, ..., m, \\ r_j &= P\{Y \leq y_j\} = \sum_{l=1}^{j} \sum_{i=1}^{m} p_{il},\ j = 1, ..., n. \end{aligned}$$

First, we prove the upper inequality in (2.1) by using Abel's method of summation (see pp. 194 of Apostol, 1974). Observe that for $u \in (0,1)$,

$$F^{-1}(u) = \sum_{i=1}^{m} x_i I_{(q_{i-1}, q_i]}(u),\ G^{-1}(u) = \sum_{j=1}^{n} y_j I_{(r_{j-1}, r_j]}(u),$$

where $q_0 = r_0 = 0$. Since $\sum_{k=1}^{i} I_{(q_{k-1}, q_k]}(u) = I_{(0, q_i]}(u)$, applying Abel's method to the expression of $F^{-1}(u)$ leads to

$$F^{-1}(u) = x_m - \sum_{i=1}^{m-1} (x_{i+1} - x_i) I_{(0, q_i]}(u),\ u \in (0,1).$$

Similarly, we have

$$G^{-1}(u) = y_n - \sum_{j=1}^{n-1}(y_{j+1} - y_j)I_{(0,r_j]}(u), \ u \in (0,1). \tag{2.2}$$

Thus,

$$\begin{aligned} E(F^{-1}(U)G^{-1}(U)) &= x_m y_n - x_m \sum_{j=1}^{n-1}(y_{j+1} - y_j)r_j - y_n \sum_{i=1}^{m-1}(x_{i+1} - x_i)q_i \\ &\quad + \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(x_{i+1} - x_i)(y_{j+1} - y_j)\min\{q_i, r_j\}. \end{aligned}$$

Also, applying Abel's method to the expressions for X and Y leads to

$$X = x_m - \sum_{i=1}^{m-1}(x_{i+1} - x_i)\sum_{k=1}^{i} I_{A_k}, \ Y = y_n - \sum_{j=1}^{n-1}(y_{j+1} - y_j)\sum_{l=1}^{j} I_{B_l},$$

from which we get

$$\begin{aligned} E(XY) &= x_m y_n - x_m \sum_{j=1}^{n-1}(y_{j+1} - y_j)r_j - y_n \sum_{i=1}^{m-1}(x_{i+1} - x_i)q_i \\ &\quad + \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(x_{i+1} - x_i)(y_{j+1} - y_j)\sum_{k=1}^{i}\sum_{l=1}^{j} p_{kl}. \end{aligned}$$

Hence, it follows that

$$\begin{aligned} &E(F^{-1}(U)G^{-1}(U)) - E(XY) \\ &\quad = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(x_{i+1} - x_i)(y_{j+1} - y_j)\Big[\min\{q_i, r_j\} - \sum_{k=1}^{i}\sum_{l=1}^{j} p_{kl}\Big] \geq 0, \end{aligned}$$

since $q_i \geq \sum_{k=1}^{i}\sum_{l=1}^{j} p_{kl}$ and $r_j \geq \sum_{k=1}^{i}\sum_{l=1}^{j} p_{kl}$.

On the other hand, from (2.2) we have

$$G^{-1}(1-u) = y_n - \sum_{j=1}^{n-1}(y_{j+1} - y_j)I_{[1-r_j,1)}(u), \ u \in (0,1),$$

and thus

$$
\begin{aligned}
E(F^{-1}(U)G^{-1}(1-U)) &= x_m y_n - x_m \sum_{j=1}^{n-1}(y_{j+1}-y_j)r_j - y_n \sum_{i=1}^{m-1}(x_{i+1}-x_i)q_i \\
&\quad + \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(x_{i+1}-x_i)(y_{j+1}-y_j)\max\{q_i+r_j-1,0\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&E(F^{-1}(U)G^{-1}(1-U)) - E(XY) \\
&= \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(x_{i+1}-x_i)(y_{j+1}-y_j)\left[\max\{q_i+r_j-1,0\} - \sum_{k=1}^{i}\sum_{l=1}^{j} p_{kl}\right] \le 0,
\end{aligned}
$$

since $q_i + r_j - \sum_{k=1}^{i}\sum_{l=1}^{j} p_{kl} \le 1$. This completes the proof. □

Theorem 2.2 *Let X and Y be random variables with distribution functions F and G, respectively. Assume that $E(X^2) < \infty$ and $E(Y^2) < \infty$. Then, we have*

$$
\begin{aligned}
-[E(X^2)E(Y^2)]^{1/2} &\le E(F^{-1}(U)G^{-1}(1-U)) \le E(XY) \\
&\le E(F^{-1}(U)G^{-1}(U)) \le [E(X^2)E(Y^2)]^{1/2},
\end{aligned}
$$

where U is a Uniform(0,1) random variable.

Proof. The lower and upper inequalities are nothing but the Cauchy-Schwarz inequalities since $F^{-1}(U) \stackrel{d}{=} X$ and $G^{-1}(U) \stackrel{d}{=} G^{-1}(1-U) \stackrel{d}{=} Y$. By $\stackrel{w}{\to}$ and $\stackrel{L^p}{\to}$ $(p > 0)$ we denote weak convergence and convergence in L^p, respectively. Since simple random variables are dense in $L^2(\Omega, \mathcal{F}, P)$, there exist simple random variables X_n and Y_n such that $X_n \stackrel{L^2}{\to} X$ and $Y_n \stackrel{L^2}{\to} Y$ as $n \to \infty$. For each n, let F_n and G_n be the distribution functions of X_n and Y_n, respectively. Then, by Lemma 2.1, for each n, we have

$$
E(F_n^{-1}(U)G_n^{-1}(1-U)) \le E(X_nY_n) \le E(F_n^{-1}(U)G_n^{-1}(U)). \tag{2.3}
$$

Since $F_n \stackrel{w}{\to} F$, $F_n^{-1}(u) \to F^{-1}(u)$ except for at most countably many u and thus $F_n^{-1}(U) \stackrel{a.s.}{\to} F^{-1}(U)$. Also, note that

$$
\lim_{n\to\infty} E(F_n^{-1}(U))^2 = \lim_{n\to\infty} E(X_n^2) = E(X^2) = E(F^{-1}(U))^2 < \infty.
$$

These imply that $\{(F_n^{-1}(U))^2\}_{n=1}^{\infty}$ is uniformly integrable and thus that $F_n^{-1}(U) \xrightarrow{L^2} F^{-1}(U)$. Similar arguments yield that $G_n^{-1}(U) \xrightarrow{L^2} G^{-1}(U)$ and $G_n^{-1}(1-U) \xrightarrow{L^2} G^{-1}(1-U)$. Therefore, we have $F_n^{-1}(U)G_n^{-1}(U) \xrightarrow{L^1} F^{-1}(U)G^{-1}(U)$ and $F_n^{-1}(U)G_n^{-1}(1-U) \xrightarrow{L^1} F^{-1}(U)G^{-1}(1-U)$, which again implies that $E(F_n^{-1}(U)G_n^{-1}(U)) \to E(F^{-1}(U)G^{-1}(U))$ and $E(F_n^{-1}(U)G_n^{-1}(1-U)) \to E(F^{-1}(U)G^{-1}(1-U))$. Likewise, $X_nY_n \xrightarrow{L^1} XY$ and so $E(X_nY_n) \to E(XY)$. Hence, letting $n \to \infty$ in (2.3), we have the desired result. $\square$

The new bounds for $E(XY)$ in Theorem 2.2 are sharper than the Cauchy-Schwarz bounds. For instance, let U_1, U_2 and U be Uniform(0,1) random variables and take $X = U_1^{\alpha}$ and $Y = U_2^{\beta}$, where $\alpha, \beta > 0$. Then $E(F^{-1}(U)G^{-1}(U)) = E(U^{\alpha+\beta}) = (\alpha+\beta+1)^{-1}$, which is less than $[E(X^2)E(Y^2)]^{1/2} = [(2\alpha+1)(2\beta+1)]^{-1/2}$ whenever $\alpha \neq \beta$.

In Theorem 2.2, it is obvious that $E(F^{-1}(U)G^{-1}(U)) = E(F^{-1}(1-U)G^{-1}(1-U))$ and $E(F^{-1}(U)G^{-1}(1-U)) = E(F^{-1}(1-U)G^{-1}(U))$ by change of variables. Since $X \overset{d}{=} F^{-1}(U)$ and $Y \overset{d}{=} G^{-1}(U) \overset{d}{=} G^{-1}(1-U)$, the corollary below follows immediately from Theorem 2.2. In particular, $F^{-1}(U)$ and $G^{-1}(U)$ are not negatively correlated while $F^{-1}(U)$ and $G^{-1}(1-U)$ are not positively correlated. This is easily verified by considering independent variables, X and Y, in Theorem 2.2.

Corollary 2.3 *Under the assumptions of Theorem 2.2, we have*

(a) $-[var(X)var(Y)]^{1/2} \leq cov(F^{-1}(U), G^{-1}(1-U)) \leq cov(X,Y)$
$\leq cov(F^{-1}(U), G^{-1}(U)) \leq [var(X)var(Y)]^{1/2}$,

(b) $cov(F^{-1}(U), G^{-1}(1-U)) \leq 0 \leq cov(F^{-1}(U), G^{-1}(U))$.

The result of Theorem 2.2 can be easily extended to the case of the Hölder inequality. The proof is similar to that of Theorem 2.2.

Theorem 2.4 *Let X and Y be random variables with distribution functions F and G, respectively. For $p > 1$ and $q > 1$ with $1/p + 1/q = 1$, assume that $E|X|^p < \infty$ and $E|Y|^q < \infty$. Then,*

$$\begin{aligned} -(E|X|^p)^{1/p}(E|Y|^q)^{1/q} &\leq E(F^{-1}(U)G^{-1}(1-U)) \leq E(XY) \\ &\leq E(F^{-1}(U)G^{-1}(U)) \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}, \end{aligned}$$

where U is a Uniform(0,1) random variable.

Theorem 2.4 informs us that $E(F^{-1}(U)G^{-1}(U))$ can be small enough to satisfy

$$E(F^{-1}(U)G^{-1}(U)) \le \inf\{(E|X|^p)^{1/p}(E|Y|^q)^{1/q} : p > 1, q > 1, 1/p + 1/q = 1\},$$

provided that X and Y have finite moments of all positive orders.

If X and Y are both nonnegative random variables, then the moment conditions in Theorems 2.2 and 2.4 are not necessary.

Theorem 2.5 *Let X and Y be nonnegative random variables with distribution functions F and G, respectively. Then,*

$$E(F^{-1}(U)G^{-1}(1-U)) \le E(XY) \le E(F^{-1}(U)G^{-1}(U)),$$

where U is a Uniform(0,1) random variable.

Proof. Let X_n and Y_n be nonnegative simple random variables such that $X_n(\omega) \uparrow X(\omega)$ and $Y_n(\omega) \uparrow Y(\omega)$ as $n \to \infty$ for all $\omega \in \Omega$. For each n, let F_n and G_n be the distribution functions of X_n and Y_n, respectively. Then, by Lemma 2.1, for each n, we have

$$E(F_n^{-1}(U)G_n^{-1}(1-U)) \le E(X_nY_n) \le E(F_n^{-1}(U)G_n^{-1}(U)).$$

Since $F_n(x) \downarrow F(x)$ for all $x \in \mathbb{R}$, $F_n^{-1}(u) \uparrow F^{-1}(u)$ for all $u \in (0,1)$ and thus $0 \le F_n^{-1}(U(\omega)) \uparrow F^{-1}(U(\omega))$ for all $\omega \in \Omega$. Similarly, we have $0 \le G_n^{-1}(U(\omega)) \uparrow G^{-1}(U(\omega))$ and $0 \le G_n^{-1}(1-U(\omega)) \uparrow G^{-1}(1-U(\omega))$ for all $\omega \in \Omega$. Thus, $0 \le F_n^{-1}(U(\omega))G_n^{-1}(U(\omega)) \uparrow F^{-1}(U(\omega))G^{-1}(U(\omega))$ and $0 \le F_n^{-1}(U(\omega))G_n^{-1}(1-U(\omega)) \uparrow F^{-1}(U(\omega))G^{-1}(1-U(\omega))$ for all $\omega \in \Omega$. By the monotone convergence theorem, we therefore have $E(F_n^{-1}(U)G_n^{-1}(U)) \to E(F^{-1}(U)G^{-1}(U))$ and $E(F_n^{-1}(U)G_n^{-1}(1-U)) \to E(F^{-1}(U)G^{-1}(1-U))$. Moreover, $0 \le X_n(\omega)Y_n(\omega) \uparrow X(\omega)Y(\omega)$ for all $\omega \in \Omega$ and so $E(X_nY_n) \to E(XY)$. Hence, the assertion follows. □

Corollary 2.3 can be used to prove monotonicity of covariance between infinitely divisible random variables. Specifically, let $\phi(t)$ be the characteristic function of an infinitely divisible distribution with finite variance. Then there exist a real number γ and a real-valued, nondecreasing, right-continuous and bounded function H defined on $\mathbb{R}$ with $H(-\infty) = \lim_{x\to-\infty} H(x) = 0$ such that

$$\phi(t) = \exp\left[i\gamma t + \int_{-\infty}^{\infty} (e^{itx} - 1 - itx)\frac{1}{x^2}\, dH(x)\right],$$

which is called the Kolmogorov representation of $\phi(t)$, and vice versa (cf. Theorem 4 in Section 12.1 of Chow and Teicher, 1988). Since, for each $\alpha > 0$, $\phi^\alpha(t)$ is also an infinitely divisible characteristic function with $\gamma^* = \alpha\gamma$ and $H^* = \alpha H$, the class of distributions with characteristic functions $\phi^\alpha(t)$, $\alpha > 0$ forms a family of one-parameter infinitely divisible distributions with finite variances. Typical examples of this family are $N(0, \alpha)$, Poisson(α), Gamma($\alpha, 1$), etc. In the following theorem we establish a monotone property for the covariance between random variables with these one-parameter infinitely divisible distribution families.

Theorem 2.6 *Let $\phi_1(t)$ and $\phi_2(t)$ be any two infinitely divisible characteristic functions with finite variances. For each $\alpha > 0$, let F_α and G_α denote the infinitely divisible distribution functions with characteristic functions $\phi_1^\alpha(t)$ and $\phi_2^\alpha(t)$, respectively. Then, for any $\alpha, \beta, \gamma, \delta > 0$, we have*

$$cov(F_\alpha^{-1}(U), G_\beta^{-1}(U)) \le cov(F_{\alpha+\gamma}^{-1}(U), G_{\beta+\delta}^{-1}(U)),$$

where U is a Uniform(0,1) random variable.

Proof. Let U_1, U_2 and U be independent Uniform(0,1) random variables. Then, from the infinitely divisible property, it follows that $F_\alpha^{-1}(U_1) + F_\gamma^{-1}(U_2) \overset{d}{=} F_{\alpha+\gamma}^{-1}(U)$ and $G_\beta^{-1}(U_1) + G_\delta^{-1}(U_2) \overset{d}{=} G_{\beta+\delta}^{-1}(U)$. Thus, by Corollary 2.3, we have

$$\begin{aligned}
&\mathrm{cov}(F_\alpha^{-1}(U_1), G_\beta^{-1}(U_1)) + \mathrm{cov}(F_\gamma^{-1}(U_2), G_\delta^{-1}(U_2)) \\
&= \mathrm{cov}(F_\alpha^{-1}(U_1) + F_\gamma^{-1}(U_2), G_\beta^{-1}(U_1) + G_\delta^{-1}(U_2)) \\
&\le \mathrm{cov}(F_{\alpha+\gamma}^{-1}(U), G_{\beta+\delta}^{-1}(U)).
\end{aligned}$$

Since $\mathrm{cov}(F_\gamma^{-1}(U_2), G_\delta^{-1}(U_2)) \ge 0$, the theorem holds. $\square$

3 Discussion

The results in Section 2 play very useful roles in constructing random vectors of correlated component variables. The bounds in Corollary 2.3 can be used to check the existence of random variables satisfying a given correlation matrix if their marginal distributions are specified. For a simple instance, suppose that $X \sim \mathrm{Uniform}(0, 1)$ and $Y \sim \mathrm{Exp}(1)$. The new upper bound for the covariance between X and Y is $1/4$ while the corresponding Cauchy-Schwarz bound is $1/(2\sqrt{3})$. Thus, the correlation coefficient should not be specified as greater than $\sqrt{3}/2$ when conducting a simulation study.

If the underlying marginal distributions are specified to be infinitely divisible, the monotone property in Theorem 2.6 is crucial in constructing their corresponding variables under specified correlation. While more general cases are discussed in Park (2002), we here consider an example of constructing a bivariate random vector (X, Y) such that $X \sim F_\alpha$, $Y \sim G_\beta$ and $\mathrm{corr}(X, Y) = \rho$, where F_α and G_β are defined as in Theorem 2.6. Let U be a Uniform(0,1) random variable and let u_α and v_β denote the variances of F_α and G_β, respectively. Then $\mathrm{cov}(X, Y) = \rho\sqrt{u_\alpha v_\beta}$ and so ρ should be specified to satisfy $\mathrm{cov}(F_\alpha^{-1}(U), G_\beta^{-1}(U)) \geq \rho\sqrt{u_\alpha v_\beta}$. If $\mathrm{cov}(F_\alpha^{-1}(U), G_\beta^{-1}(U)) = \rho\sqrt{u_\alpha v_\beta}$, then just take $X = F_\alpha^{-1}(U)$ and $Y = G_\beta^{-1}(U)$. Otherwise, there must exist a $\gamma \in (0, \beta)$ such that

$$\mathrm{cov}(F_\alpha^{-1}(U), G_\gamma^{-1}(U)) = \rho\sqrt{u_\alpha v_\beta}, \tag{3.1}$$

and so by taking $X = F_\alpha^{-1}(U_1)$ and $Y = G_\gamma^{-1}(U_1) + G_{\beta-\gamma}^{-1}(U_2)$ with U_1 and U_2 being independent Uniform(0,1) random variables we can construct the desired random vector. A searching algorithm for the value of γ in (3.1) can be easily developed from the monotone property in Theorem 2.6.

REFERENCES

Apostol, T. M. (1974). *Mathematical Analysis.* Addison Wesley.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory.* (2nd ed.). Springer-Verlag, New York.

Feagan, B. G., Rochon, J., Fedorak, R. N., Irvine, E. J., Wild, G., Sutherlan, L., Steinhart, A. H., Greenberg, G. R., Gilles, R., Hopkins, M., Hanauer, S. B. and McDonald, J. W. D. (1995). Methotrexate for treatment of Crohn's disease. *New England Journal of Medicine* **332**, 292-297.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **44**, 13-22.

Park, C. G. (2003). Construction of random vectors of heterogeneous component variables under specified correlation structures. Computational Statistics and Data Analysis (to appear).

Regan, M. M. and Catalano, P. J. (1999). Likelihood models for clustered binary and continuous outcomes: application to developmental toxicology. *Biometrics* **55**, 760-768.

Rochon, J. (1996). Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics* **52**, 760-768.

Zeger, S. L, Liang, K. Y. and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equations approach. *Biometrics* **44**, 1049-1060.