Research Article
A New Generalized Inequality for Covariance in \( N \) Dimensions

Mei Liu,\(^1\) Milan Šindelka,\(^2\) and Shi-you Lin\(^3\)

\(^1\)School of Mathematics and Statistics, Hainan Normal University, Haikou, Hainan, 571158, China
\(^2\)Institute of Plasma Physics, Czech Academy of Sciences, Za Slovankou 3, 18200 Prague 8, Czech Republic

Correspondence should be addressed to Shi-you Lin; linsy1111@foxmail.com

Received 19 August 2018; Revised 15 December 2018; Accepted 24 December 2018; Published 16 January 2019

1. Introduction

The concept of covariance appears ubiquitously in probability theory and statistics as the basic measure of correlation between random variables (see, e.g., standard textbooks [1–5]). An intuitive though elusive idea of correlation was transformed into a sound mathematical language in the 19\(^{th}\) century by Auguste Bravais and Francis Galton. Readers interested in the fascinating history of this subject are advised to consult a comprehensive work of Stingler [6] or two tutorial articles [7, 8]. An analysis of covariance possesses great practical importance in applied sciences [9, 10], especially in engineering (error analysis, optimum control, probabilistic design, and system identification) [11, 12], in biotechnology [13] and medical sciences [14, 15], or in economy [16].

Naturally, the means/variances/covariances of any ensemble of random variables are uniquely determined by the probability distribution (PD) corresponding to the concrete problem under study. However, both in pure mathematics and in applications, one frequently encounters a case when the pertinent PD is unknown. Under such circumstances, it is often much desirable to provide at least well defined general estimates (constraints) regarding mean/variance/covariance, namely, such estimates which are independent upon the specific PD. These constraints take typically the form of an inequality.

One notable result of the abovementioned sort was obtained by Chebyshev as early as in 1882 [1–3, 6]. More explicitly, the so-called Chebyshev inequality enables one to estimate which maximum fraction of values of a given random variable can be located further than a prescribed distance from the mean. Closely related are the subsequently found Ostrowski and Grüss type inequalities [1–3] in all their different variants (as listed in comprehensive monographs [17–19]).

Until nowadays, the works of Chebyshev, Ostrowski, and Grüss have continued to inspire active mathematical research focused on inequalities/estimates of (co)variance. This fact is clearly documented by rich literature dealing with the subject [20–37]. For the purposes of the present article we specifically highlight an inequality for covariance which was derived recently by He and Wang in [20]. Namely, the following theorem has been proven to hold.

**Theorem I.**

\[
|\text{cov}(f(\xi), g(\xi))| \leq 2 \|f^'\|_\infty \|g^'\|_\infty D\xi. \tag{1}
\]

Here \( \xi \in (a, b) \) is a single random variable. One assumes that \( \xi \) has a finite expectation value \( E\xi \) and a finite variance...
### 2. Preparatory Considerations

#### 2.1. Multidimensional Random Variables

##### 2.1.1. Discrete Case. Assume that

(i) \( \mathcal{M} \) is a countable set, \( \mathcal{M} \subseteq \mathbb{R}^N \)

(ii) \( \xi = (\xi_1, \ldots, \xi_N) \) is an \( N \)-dimensional random variable, \( \xi \in \mathcal{M} \)

(iii) \( \varphi(\xi) \geq 0 \) (\( \forall \xi \in \mathcal{M} \)) are the associated probabilities completely determining the statistical distribution of \( \xi \)

Then

(a) the probabilities \( \varphi(\xi) \) satisfy the normalization condition

\[
\sum_{\xi \in \mathcal{M}} \varphi(\xi) = 1
\]

(b) an expectation of \( \xi \) is given by formula

\[
E\xi = \sum_{\xi \in \mathcal{M}} \xi \varphi(\xi).
\]

Hereafter we shall assume that \( E\xi \) is well defined and finite

(c) the variance of \( \xi_j \) (with \( j \in \{1, \ldots, N\} \)) is given by formula

\[
D\xi_j = E\left(\xi_j - E\xi_j\right)^2.
\]

##### 2.1.2. Continuous Case. Assume that

(i) \( \Omega \) is an open set, \( \Omega \subseteq \mathbb{R}^N \)

(ii) \( \xi = (\xi_1, \ldots, \xi_N) \) is an \( N \)-dimensional random variable, \( \xi \in \Omega \)

(iii) \( \varphi(\xi) \geq 0 \) (\( \forall \xi \in \Omega \)) are the associated probability densities completely determining the statistical distribution of \( \xi \)

Then

(a) the probability densities \( \varphi(\xi) \) satisfy the normalization condition

\[
\int_{\Omega} \varphi(\xi) \, d^N\xi = 1
\]
(b) an expectation of $\xi$ is given by formula
\[
E\xi = \int_{\Omega} \xi \varphi(\xi) d^N \xi.
\]  
(13)

Hereafter we shall assume that $E\xi$ is well defined and finite

(c) the variance of $\xi_j$ (with $j \in \{1, \ldots, N\}$) is given by formula
\[
D\xi_j = E(\xi_j - E\xi_j)^2.
\]  
(14)

Hereafter we shall assume that $D\xi_j$ is well defined and finite. Similarly as above in (5), also relation (14) can be recast into an equivalent form
\[
D\xi_j = E\xi_j^2 - E\xi_j^2.
\]  
(15)

Let $f : \Omega \to \mathbb{R}$ be a function which is measurable and bounded in $\Omega$. Boundedness of $f$ means that
\[
\|f\|_\infty = \sup_{\xi \in \Omega} |f(\xi)| < \infty.
\]  
(16)

Then

(a) an expectation
\[
Ef(\xi) = \int_{\Omega} f(\xi) \varphi(\xi) d^N \xi
\]  
(17)
is finite and well defined, since
\[
|Ef(\xi)| \leq \int_{\Omega} |f(\xi)| \varphi(\xi) d^N \xi \leq \|f\|_\infty \int_{\Omega} \varphi(\xi) d^N \xi
\]  
(18)

\[
= \|f\|_\infty < \infty
\]

(b) also the variance
\[
Df(\xi) = E(f(\xi) - Ef(\xi))^2 = Ef^2(\xi) - (Ef(\xi))^2
\]  
(19)
is finite and well defined, since
\[
Ef^2(\xi) = \int_{\Omega} f^2(\xi) \varphi(\xi) d^N \xi \leq \|f\|_\infty^2 \int_{\Omega} \varphi(\xi) d^N \xi
\]  
(20)

\[
= \|f\|_\infty^2 < \infty
\]

2.2. Covariance. Let $\xi = (\xi_1, \ldots, \xi_N)$ be an $N$-dimensional random variable. We define the covariance $\text{cov}(\xi_j, \xi_j)$ by prescription
\[
\text{cov}(\xi_j, \xi_j) = E(\xi_j - E\xi_j)(\xi_j - E\xi_j),
\]  
(21)

with $j_1, j_2 \in \{1, 2, \ldots, N\}$. Equivalently one can write
\[
\text{cov}(\xi_j, \xi_j)
\]  
(22)

Recall that $\text{cov}(\xi_j, \xi_j) = D\xi_j$ as follows from (5)-(6). After taking an absolute value of (21) one gets
\[
|\text{cov}(\xi_j, \xi_j)| \leq E|\xi_j - E\xi_j| |\xi_j - E\xi_j|.
\]  
(23)

The well known Cauchy-Schwarz inequality (to be reviewed in the following subsection) implies however
\[
E|\xi_j - E\xi_j| |\xi_j - E\xi_j| \leq \sqrt{E(\xi_j - E\xi_j)^2} \sqrt{E(\xi_j - E\xi_j)^2}.
\]  
(24)

Hence one may conclude that
\[
|\text{cov}(\xi_j, \xi_j)| \leq \sqrt{D\xi_j D\xi_j},
\]  
(25)

This means also that $\text{cov}(\xi_1, \xi_j)$ is well defined and finite as long as all the variances $D\xi_j$ are well defined and finite.

2.3. The Cauchy-Schwartz Inequality. Let $x = (x_1, \ldots, x_N)$ and $y = (y_1, \ldots, y_N)$ be any two vectors in $\mathbb{R}^N$. Then
\[
\left( \sum_{j=1}^N x_j y_j \right)^2 \leq \left( \sum_{j=1}^N x_j^2 \right) \left( \sum_{j=1}^N y_j^2 \right).
\]  
(26)

The proof is discussed in all standard textbooks of functional analysis, e.g., in [38].

2.4. The Lagrange Mean Value Theorem in N Dimensions. Let $f$, $g$ be real valued differentiable functions defined in an open convex set $\Omega \subseteq \mathbb{R}^N$. Let $\xi = (\xi_1, \ldots, \xi_N) \in \Omega$, $\eta = (\eta_1, \ldots, \eta_N) \in \Omega$. Convexity of $\Omega$ implies that a straight line segment $[\xi + \gamma(\eta - \xi)]$, $\gamma \in [0, 1]$ connecting $\xi$ with $\eta$ is entirely contained within $\Omega$. The Lagrange mean value theorem states that there always exists a number $\gamma(\xi, \eta) \in (0, 1)$ with the basic property
\[
f(\eta) - f(\xi) = \sum_{j=1}^N \partial_j f(\xi + \gamma(\eta - \xi))(\eta_j - \xi_j).
\]  
(27)

Because it is not easy to find a proof of statement (27) in available standard textbooks, we prefer to supplement here our own short proof. It is inspired by Theorem 4.2 given on page 378 of [39]: Direct calculation confirms that
\[
\int_0^1 \left\{ \sum_{j=1}^N \partial_j f(\xi + \gamma(\eta - \xi))(\eta_j - \xi_j) \right\} d\gamma
\]  
(28)

\[
= \int_0^1 \frac{d}{d\gamma} f(\xi + \gamma(\eta - \xi)) d\gamma = f(\eta) - f(\xi).
\]

The Lagrange mean value theorem of the integral calculus (see, e.g., [40]) implies however that
\[
\frac{d}{d\gamma} f(\xi + \gamma(\eta - \xi)) |_{\gamma = \gamma(\xi, \eta)} = \sum_{j=1}^N \partial_j f(\xi + \gamma(\eta - \xi))(\eta_j - \xi_j).
\]  
(29)
where \( \gamma(\xi, \eta) \in (0, 1) \) is a generally unknown fixed number (depending of course not only upon \( \xi \) and \( \eta \) but also on the function \( f \)). Combination of (28) and (29) yields now immediately the desired claim (27).

### 3. Multidimensional Generalization of Theorem I

#### 3.1. Preliminaries

Before formulating the above advertised multidimensional generalization of Theorem I, let us introduce some additional notations and conventions. Recall that \( \xi \) is an \( N \)-dimensional random variable with a finite expectation and a finite variance (see Section 2.1 for details). We define an auxiliary quantity

\[
D\xi = \sum_{j=1}^{N} D\xi_j. \tag{30}
\]

Symbol \( \Omega \) will hereafter stand for an open convex subset of \( \mathbb{R}^N \) (case \( \Omega = \mathbb{R}^N \) is also allowed to occur). If \( \xi \) is a discrete random vector, then we tacitly assume \( \mathcal{M} \subset \Omega \).

Assume now that two functions \( f, g : \Omega \to \mathbb{R} \) are continuous and differentiable in \( \Omega \). Assume also that all the partial derivatives \( \partial_j f \), \( \partial_j g \) are bounded in \( \Omega \). Then one may define auxiliary symbols

\[
\|f\|_\infty = \sup_{\xi \in \Omega} |\partial_j f (\xi)| < \infty, \\
\|g\|_\infty = \sup_{\xi \in \Omega} |\partial_j g (\xi)| < \infty. \tag{31}
\]

(1 \( \leq j \leq N \)).

Subsequently one may introduce additional notations \( f_\infty \) and \( g_\infty \) through the formulas

\[
f_\infty = \sqrt{\sum_{j=1}^{N} \|f\|_\infty^2}, \\
g_\infty = \sqrt{\sum_{j=1}^{N} \|g\|_\infty^2}. \tag{32}
\]

#### 3.2. Our Basic Theorem

Now we are ready to state our own multidimensional generalization of Theorem I.

**Theorem II.** Assuming the above specified notations and conventions, one has

\[
|\text{cov} (f(\xi), g(\xi))| \leq 2f_\infty g_\infty D\xi. \tag{33}
\]

**Proof.** Since the proof is a bit lengthy, we shall conveniently divide it into parts:

(i) let \( x \in \Omega \) be a fixed parameter. Then \( f(x) \) is just a fixed number, and

\[
f(x) - Ef(\xi) = E(f(x) - f(\xi)). \tag{34}
\]

Define a quantity

\[
Q(x) = (f(x) - Ef(\xi))^2 = E^2(f(x) - f(\xi)). \tag{35}
\]

The Lagrange mean value theorem (Section 2.4) states that

\[
f(x) - f(\xi) = \sum_{j=1}^{N} \partial_j f (\xi + \gamma (\xi, x)(x - \xi)) (x_j - \xi_j), \tag{36}
\]

where \( 0 < \gamma(\xi, x) < 1 \). Hence

\[
Q(x) = E \left\{ \sum_{j=1}^{N} \partial_j f (\xi + \gamma (\xi, x)(x - \xi)) (x_j - \xi_j) \right\}^2 \leq \left( \sum_{j=1}^{N} E \left\{ \partial_j f (\xi + \gamma (\xi, x)(x - \xi)) (x_j - \xi_j) \right\} \right)^2. \tag{37}
\]

In the last line of (37) we have used the triangle inequality

(ii) by definition (17), we have

\[
E \left\{ \partial_j f (\xi + \gamma (\xi, x)(x - \xi)) (x_j - \xi_j) \right\} = \int_{\Omega} \partial_j f (\xi + \gamma (\xi, x)(x - \xi))(x_j - \xi_j) \varphi(\xi) \, d^N\xi. \tag{38}
\]

But the involved integral satisfies an inequality

\[
\int_{\Omega} \partial_j f (\xi + \gamma (\xi, x)(x - \xi))(x_j - \xi_j) \varphi(\xi) \, d^N\xi \leq \int_{\Omega} \left| \partial_j f (\xi + \gamma (\xi, x)(x - \xi)) \right| |x_j - \xi_j| \varphi(\xi) \, d^N\xi \\
\leq \|f\|_\infty \int_{\mathbb{R}^N} |x_j - \xi_j| \varphi(\xi) \, d^N\xi = \|f\|_\infty E|x_j - \xi_j|. \tag{39}
\]

Relations (38)-(39) correspond of course to continuous statistical distributions. Analogical formulas apply also in the case of discrete random variables (one merely replaces \( \int_{\Omega} \, d^N\xi \) by \( \sum_{\xi \in \mathcal{M}} \)). We leave the details to the reader

(iii) after plugging (39) into (37) one finds that

\[
Q(x) \leq \left( \sum_{j=1}^{N} \|f\|_\infty E|x_j - \xi_j| \right)^2. \tag{40}
\]
Cauchy-Schwarz inequality (26) implies however
\[ \left( \sum_{j=1}^{N} \| f_j \|_{\infty}^2 \right)^{1/2} \leq \sum_{j=1}^{N} \| f_j \|_{\infty} \sum_{j=1}^{N} \| x_j - \xi_j \|_1. \] (41)

Hence also
\[ Q(x) \leq \ell_{\infty}^{2} \sum_{j=1}^{N} \| x_j - \xi_j \|_1, \] (42)

where
\[ \ell_{\infty} = \sqrt{\sum_{j=1}^{N} \| f_j \|_{\infty}^2}. \] (43)

consonantly with (32)
(iv) proceeding further, the variance relation
\[ E[x_j - \xi_j^2] - E^2[x_j - \xi_j] = D[x_j - \xi_j] \geq 0 \] (44)
implies \( E^2[x_j - \xi_j] \leq E(x_j - \xi_j)^2, \) giving in turn
\[ Q(x) \leq \ell_{\infty}^{2} \sum_{j=1}^{N} \left( x_j - \xi_j \right)^{2}. \] (45)

Yet the variance formula
\[ E\left( x_j - \xi_j \right)^2 - E^2\left( x_j - \xi_j \right) = D\left( x_j - \xi_j \right) \] (46)
yields \( E(x_j - \xi_j)^2 = E^2(x_j - \xi_j) + D(x_j - \xi_j) = (x_j - E\xi_j)^2 + D\xi_j. \)
This is valid since \( Dx_j = 0 \) and \( D(-\xi_j) = D\xi_j. \) We may thus conclude that
\[ Q(x) \leq \ell_{\infty}^{2} \sum_{j=1}^{N} \left[ (x_j - E\xi_j)^2 + D\xi_j \right]. \] (47)

(v) comparison of (35) and (47) provides an inequality
\[ (f(x) - Ef(\xi)) \leq \ell_{\infty}^{2} \sum_{j=1}^{N} \left[ (x_j - E\xi_j)^2 + D\xi_j \right]. \] (48)

So far, an entity \( x \) was treated as a fixed parameter
(vi) now we shall set \( x \) to be a random variable equivalent to \( \xi. \) Consequently we take the expectation over (48). This results in
\[ E(f(\xi) - Ef(\xi))^2 \leq \ell_{\infty}^{2} \sum_{j=1}^{N} \left[ E(\xi_j - E\xi_j)^2 + D\xi_j \right]. \] (49)

Yet, by the definition of the variance (19), one can convert (49) into a simple outcome
\[ Df(\xi) \leq \ell_{\infty}^{2} \cdot 2D\xi, \] (50)

where of course
\[ D\xi = \sum_{j=1}^{N} D\xi_j, \] (51)

consonantly with (30)
(vii) completely analogous sequence of considerations can be applied also to the case of \( g(\xi). \) One would arrive towards an inequality
\[ Dg(\xi) \leq \varpi_{\infty} \cdot 2D\xi, \] (52)

where
\[ \varpi_{\infty} = \sqrt{\sum_{j=1}^{N} \| g_j \|_{\infty}^2}. \] (53)

Combination of (25), (50), and (52) provides now the desired final claim
\[ |\text{cov}(f(\xi), g(\xi))| \leq \sqrt{Df(\xi)Dg(\xi)} \leq 2\varpi_{\infty} \cdot \varpi_{\infty} \cdot D\xi. \] (54)

This is exactly as stated above in (33). Thus, our Theorem II is proven.

4. An Application of Theorem II on Different Probability Distributions

In the present section, we shall derive some new inequalities by applying our basic Theorem II to three different types of probability distributions of an \( N \)-dimensional random variable. The definitions of the distributions discussed below can be found in [4, 41].

4.1. Multinomial Distribution

4.1.1. Definition. The probability density function \( \varphi(\xi) \) is given by formula
\[ \varphi(\xi) = \begin{cases} \beta & \text{if } \xi \in \Omega, \\ 0 & \text{if otherwise.} \end{cases} \] (55)

Notice that
(a) the probability densities \( \varphi(\xi) \) satisfy the normalization condition
\[ \int_{\Omega} \varphi(\xi) d^N\xi = 1. \] (56)

Therefore
\[ \beta \int_{\Omega} d^N\xi = 1. \] (57)
Define
\[ V_\Omega = \int_\Omega d^N \xi; \]  
(58)

this is the volume of region \( \Omega \). Then
\[ \beta = \frac{1}{V_\Omega} \]  
(59)

(b) the expectation of \( \xi_j \) (with \( j \in \{1, \ldots, N\} \)) is given by formula combining (55), (59), and (13),
\[ \mathbb{E} \xi_j = \frac{1}{V_\Omega} \int_\Omega \xi_j d^N \xi \]  
(60)

(c) the variance of \( \xi_j \) (with \( j \in \{1, \ldots, N\} \)) is then given by formula combining (15) and (60),
\[ \text{Def} \]  
(61)

Therefore
\[ \text{Def} \]  
(62)

4.1.2. An Application of Theorem II

\textbf{Theorem II-1.} Assume that two functions \( f, g : \Omega \rightarrow \mathbb{R} \) are continuous and differentiable in \( \Omega \). Assume also that all the partial derivatives \( \partial f / \partial x_j, \partial g / \partial x_j \) are bounded in \( \Omega \). Recalling notations (31) and (32), one has
\[ |\mathbb{E} (f, g)| \]  
6 Mathematical Problems in Engineering

and therefore
\[ \text{Def} \]  
(63)

where
\[ \mathbb{E} \]  
(64)

Proof. Let \( \xi \) be an \( N \)-dimensional random variable which possesses the multivariate uniform distribution. According to (55), (59), and (17) we have
\[ \text{Def} \]  
(65)

Subsequently we have
\[ D_\xi = \sum_{j=1}^{N} \sum_{j=1}^{N} (b_j - a_j)^2 \]  

(74)

Now we are ready to make the following statement.

**Theorem II-2.** One has

\[ |\mathcal{F}(f, g)| \leq \frac{1}{6} \sum_{j=1}^{n} (b_j - a_j)^2 \mathcal{I}_{\infty} g_{\infty}, \]

where

\[ \mathcal{F}(f, g) = \prod_{j=1}^{N} (b_j - a_j) \int_{a_j}^{b_j} f(\xi) g(\xi) \, d\xi_1 \ldots d\xi_N \]

(75)

- (ii)

\[ \text{Theorem II-3. Assume that two functions } f, g : \mathcal{M} \to \mathbb{R} \]

are continuous and differentiable in \( \Omega \) where \( \mathcal{M} \subset \Omega \subseteq \mathbb{R}^N \)

Assume also that all the partial derivatives \( \{ \partial_i f \}_{i=1}^{N}, \{ \partial_i g \}_{i=1}^{N} \)

are bounded in \( \Omega \). Recalling notations (31) and (32), one has

\[ \lim_{k \to \infty} \sum_{j=1}^{n} n! \prod_{j=0}^{N} \frac{p_j}{p_j} \frac{p_j}{p_j} \ldots \frac{p_j}{p_j} \]

(81)

\[ \leq 2 \mathcal{I}_{\infty} g_{\infty} \sum_{j=1}^{N} p_j \]

Proof. Let \( \xi \) be an \( N \)-dimensional random variable which possesses the multinomial distribution. According to (78) and (8) we have

\[ Ef(\xi) = \sum_{\xi=0}^{N} \sum_{\xi=0}^{N} \sum_{\xi=0}^{N} f(\xi) \]

(82)

Therefore

\[ Ef(\xi) g(\xi) - Ef(\xi) Eg(\xi) \]

\[ = \sum_{\xi=0}^{N} \sum_{\xi=0}^{N} \sum_{\xi=0}^{N} f(\xi) g(\xi) \]

\[ \leq \sum_{\xi=0}^{N} \sum_{\xi=0}^{N} \sum_{\xi=0}^{N} \frac{n!}{\xi_1! \xi_2! \ldots \xi_N!} p_1^{e_1} p_2^{e_2} \ldots p_N^{e_N} \]

(83)

\[ \leq 2 \mathcal{I}_{\infty} g_{\infty}, D_{\xi}. \]

Combination of (83), (84), and (80) yields desired inequality (81); thus the proof is completed. \( \square \)
4.3. One-Dimensional Normal Distribution. Since the normal distribution was not discussed in paper [20] of He and Wang, we shall discuss first the case $N = 1$ of a single random variable. Later in the subsequent subsection we shall extend our result to the case of two random variables ($N = 2$).

4.3.1. Definition. Assume here

(i) $\Omega = (-\infty, +\infty)$

(ii) $\xi \in \Omega$ is a one-dimensional random variable

(iii) $\mu, \sigma > 0$ are given prescribed parameters

Then

(a) the probability density function $\varphi(\xi)$ is given by formula

$$\varphi(\xi) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\xi - \mu)^2}{2\sigma^2}}$$  \hspace{1cm} (85)

(b) the expectation is equal to

$$E\xi = \mu$$  \hspace{1cm} (86)

(c) the variance is equal to

$$D\xi = \sigma^2$$  \hspace{1cm} (87)

4.3.2. An Application of Theorem I

Theorem I-1. Assume that two functions $f, g : \Omega \rightarrow \mathbb{R}$ are continuous and differentiable in $\Omega$. Assume also that the derivatives $f', g'$ are bounded in $\Omega$. Then

$$\left| \int_{-\infty}^{+\infty} f(\xi) g(\xi) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\xi - \mu)^2}{2\sigma^2}} d\xi \right|$$

$$\leq 2 \left\| f' \right\|_\infty \left\| g' \right\|_\infty$$  \hspace{1cm} (88)

Proof. Let $\xi$ be a random variable which possesses the normal distribution. According to (85) and (17) we have

$$Ef(\xi) = \int_{-\infty}^{+\infty} f(\xi) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\xi - \mu)^2}{2\sigma^2}} d\xi;$$  \hspace{1cm} (89)

therefore

$$Ef(\xi) g(\xi) - Ef(\xi) Eg(\xi)$$

$$= \int_{-\infty}^{+\infty} f(\xi) g(\xi) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\xi - \mu)^2}{2\sigma^2}} d\xi$$

$$- \int_{-\infty}^{+\infty} f(\xi) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\xi - \mu)^2}{2\sigma^2}} d\xi \cdot g(\xi)$$  \hspace{1cm} (90)

Yet (1) of Theorem I implies

$$\left| Ef(\xi) g(\xi) - Ef(\xi) Eg(\xi) \right| = \left| \text{cov} (f(\xi), g(\xi)) \right|$$

$$\leq 2\int_{-\infty}^{+\infty} f(\xi) g(\xi) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\xi - \mu)^2}{2\sigma^2}} d\xi$$  \hspace{1cm} (91)

Combination of (90), (91), and (87) yields desired inequality (88); thus the proof is completed.

4.3.3. Standard Normal Distribution. The standard normal distribution corresponds to choosing $\mu = 0$ and $\sigma = 1$. Accordingly we have

$$\varphi(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}.$$  \hspace{1cm} (92)

After substituting (92) into (88) one arrives to the following outcome.

Theorem I-2.

$$\left| \int_{-\infty}^{+\infty} f(\xi) g(\xi) \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi \right|$$

$$- \int_{-\infty}^{+\infty} f(\xi) \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi \int_{-\infty}^{+\infty} g(\xi) \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi$$  \hspace{1cm} (93)

$$\leq 2 \left\| f' \right\|_\infty \left\| g' \right\|_\infty.$$  

4.4. Standard Bivariate Normal Distribution

4.4.1. Definition. Assume that

(i) $\Omega = \mathbb{R}^2$

(ii) $\xi = (\xi_1, \xi_2)$ with $\xi_1$ and $\xi_2$ independent

(iii) $\phi(\xi_j)$ denotes the standard normal probability density function of $\xi_j$, with $j \in \{1, 2\}$:

$$\phi(\xi_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi_j^2}{2}}$$  \hspace{1cm} (94)

Then

(a) the probability density function $\varphi(\xi)$ is given by formula

$$\varphi(\xi) = \phi(\xi_1) \phi(\xi_2) = \frac{1}{2\pi} e^{-\frac{(\xi_1^2 + \xi_2^2)}{2}}$$  \hspace{1cm} (95)

(b) the expectation is equal to

$$E\xi = 0$$  \hspace{1cm} (96)

(c) the variance is equal to

$$D\xi = D\xi_1 + D\xi_2 = 2$$  \hspace{1cm} (97)
4.4.2. An Application of Theorem II

**Theorem II-4.** Assume that two functions \( f, g : \Omega \rightarrow \mathbb{R} \) are continuous and differentiable in \( \Omega \). Assume also that all the partial derivatives \( \partial f_N^{(j)} \), \( \partial g_N^{(j)} \) are bounded in \( \Omega \). Recalling (31) and (32), one has

\[
\left| \int f(\xi)\,g(\xi) - \int f(\xi)\,\text{cov}(f(\xi), g(\xi)) \right| \leq 2f_{\text{co}}g_{\text{co}}D_{\xi}.
\]

Proof. Let \( \xi \) be a 2-dimensional random variable which possesses the standard bivariate normal distribution. According to (95) and (17), one has

\[
E_f(\xi) = \int f(\xi) \frac{1}{2\pi} e^{-\frac{(\xi_1^2 + \xi_2^2)}{2}} \, d\xi_1 \, d\xi_2;
\]

therefore

\[
E_f(\xi)g(\xi) - Ef(\xi)Eg(\xi) = \left| \frac{1}{2\pi} e^{-\frac{(\xi_1^2 + \xi_2^2)}{2}} \int \xi_1 d\xi_1 \, d\xi_2 \right| 
\]

\[
\leq 2f_{\text{co}}g_{\text{co}}D_{\xi}.
\]

Combination of (100), (101), and (97) yields desired inequality (98); thus the proof is completed.

5. Coordinate Dependence of Theorem II

Our basic relation (33) of Theorem II is formulated in terms of an \( N \)-dimensional random variable \( \xi \in \Omega \subseteq \mathbb{R}^N \). Clearly, for a given \( N \)-dimensional statistical problem there are many different (mutually equivalent) choices of \( N \) independent random variables (coordinates) \( \xi \) in terms of which the probability distribution \( \varphi(\xi) \) can be expressed. It is not a priori clear whether or not our basic random inequality (33) is coordinate dependent. As we show below on a concrete example, an answer is affirmative. This means also that an inequality (33) can be optimized (made stronger) via carrying out a suitable coordinate transformation.

5.1. Correlated Bivariate Normal Distribution. As an illustrative example we shall take the correlated bivariate normal distribution described in [4]. Correspondingly, we have \( N = 2, \Omega = \mathbb{R}^2 \), and \( \xi = (x, y) \). The probability distribution of \( \xi \) is characterized by formula

\[
\varphi(\xi) = \frac{1}{2\pi \sqrt{1 - \rho^2}} e^{-\frac{(x^2 - 2\rho xy + y^2)}{2(1 - \rho^2)}},
\]

where \( \rho \in [0, 1] \) is a fixed correlation parameter.

Importantly, the above introduced correlated bivariate normal distribution can be converted into an uncorrelated normal distribution via performing a bijective coordinate transformation \( \xi = (x, y) \leftrightarrow \eta = (x, z) \). Here, according to [4],

\[
z = \frac{y - \rho x}{\sqrt{1 - \rho^2}};
\]

\[
y = \rho x + \sqrt{1 - \rho^2}z.
\]

The associated probability density \( \varphi(\eta) \) takes the form already discussed above in (95), i.e.,

\[
\varphi(\eta) = \frac{1}{2\pi} e^{-\frac{(z^2 + x^2)}{2}}.
\]

Direct calculation yields \( D_x = D_z = 1 \); thus \( D_\eta = D_x + D_z = 1 + 1 = 2 \). Hence also

\[
D_\xi = D_x + D_y = D_x + D \left\{ \rho x + \sqrt{1 - \rho^2 z} \right\} = D_x + \rho^2 D_x + (1 - \rho^2) D_z = 2.
\]

5.2. An Application of Theorem II: Correlated Case. For the sake of maximum simplicity, we shall apply Theorem II on a concrete case of two elementary functions

\[
f(x, y) = x + y;
\]

\[
g(x, y) = x - y.
\]

Direct calculation yields

\[
\|f_x\| = \sup_{(x,y) \in \mathbb{R}^2} \frac{|\partial f|}{|\partial x|} = 1;
\]

\[
\|g_x\| = \sup_{(x,y) \in \mathbb{R}^2} \frac{|\partial g|}{|\partial x|} = 1;
\]

\[
\|f_y\| = \sup_{(x,y) \in \mathbb{R}^2} \frac{|\partial f|}{|\partial y|} = 1;
\]

\[
\|g_y\| = \sup_{(x,y) \in \mathbb{R}^2} \frac{|\partial g|}{|\partial y|} = 1;
\]

\[
\|f_{\text{co}}\| = \sqrt{\|f_x\|^2 + \|f_y\|^2} = \sqrt{2},
\]

\[
\|g_{\text{co}}\| = \sqrt{\|g_x\|^2 + \|g_y\|^2} = \sqrt{2}.
\]
Relation (33) of Theorem II boils then down to

$$|\text{cov} (f(\xi), g(\xi))| \leq 8,$$  \hspace{1cm} (110)

where

$$\text{cov} (f(\xi), g(\xi)) = E_f(\xi)g(\xi) - \text{E}_f(\xi)\text{E}g(\xi)$$

$$= \iint f(x, y)g(x, y) \cdot \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/(2(1-\rho^2))} \, dx \, dy$$

$$- \iint f(x, y) \cdot \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/(2(1-\rho^2))} \, dx \, dy \iint g(x, y) \cdot \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/(2(1-\rho^2))} \, dx \, dy.$$  \hspace{1cm} (111)

5.3. An Application of Theorem II: Uncorrelated Case. In the coordinates $\eta = (x, z)$ introduced above by (104) we have

$$F(x, z) = f(x, y(x, z)) = x + \left(\rho x + \sqrt{1-\rho^2}z\right)$$

$$= (1+\rho)x + \sqrt{1-\rho^2}z;$$

$$G(x, y) = g(x, y(x, z)) = x - \left(\rho x + \sqrt{1-\rho^2}z\right)$$

$$= (1-\rho)x - \sqrt{1-\rho^2}z.$$  \hspace{1cm} (112)

Direct calculation yields

$$\left\|F_x\right\| = \sup_{(x, z) \in \mathbb{R}^1} \left|\frac{\partial F}{\partial x}\right| = 1 + \rho,$$

$$\left\|G_x\right\| = \sup_{(x, z) \in \mathbb{R}^1} \left|\frac{\partial G}{\partial x}\right| = 1 - \rho;$$

$$\left\|F_z\right\| = \sup_{(x, z) \in \mathbb{R}^1} \left|\frac{\partial F}{\partial z}\right| = \sqrt{1-\rho^2},$$

$$\left\|G_z\right\| = \sup_{(x, z) \in \mathbb{R}^1} \left|\frac{\partial G}{\partial z}\right| = \sqrt{1-\rho^2};$$

$$\mathcal{G}_{\infty} = \sqrt{\left\|F_x\right\|^2 + \left\|F_z\right\|^2} = \sqrt{2}\sqrt{1+\rho},$$

$$\mathcal{G}_{\infty} = \sqrt{\left\|G_x\right\|^2 + \left\|G_z\right\|^2} = \sqrt{2}\sqrt{1-\rho}.$$  \hspace{1cm} (114)

According to our above derived statement (98) of Theorem II-4, relation (33) of Theorem II boils down to

$$|\text{cov} (F(\eta), G(\eta))| \leq 8\sqrt{1-\rho^2},$$  \hspace{1cm} (115)

where

$$\text{cov} (F(\eta), G(\eta)) = E_f(\xi)g(\xi) - \text{E}_f(\xi)\text{E}g(\xi)$$

$$= \iint F(x, z)G(x, z) \frac{1}{2\pi} e^{(x^2+z^2)/2} \, dx \, dz$$

$$- \iint F(x, z) \frac{1}{2\pi} e^{(x^2+z^2)/2} \, dx \, dz$$

$$\cdot \iint G(x, z) \frac{1}{2\pi} e^{(x^2+z^2)/2} \, dx \, dz.$$

5.4. Analysis. Functions $f(x, y)$ and $F(x, z)$ represent the same random variable, just expressed in different coordinates. This is similar for $g(x, y)$ and $G(x, z)$. If so, then covariances (111) and (116) must be equal. This can easily be verified explicitly as a consistency check. One starts with (111) and performs a substitution $(x, y) \rightarrow (x, z)$ following (104). The corresponding Jacobian is equal to

$$f(x, z) = \frac{\partial x}{\partial x} \frac{\partial y}{\partial y} = \left|\begin{array}{cc} 1 & \rho \\ 0 & 1-\rho^2 \end{array}\right| = \sqrt{1-\rho^2}.$$  \hspace{1cm} (117)

Straightforward manipulations confirm then that

$$\text{cov} (f(\xi), g(\xi)) = \text{cov} (F(\eta), G(\eta)),$$  \hspace{1cm} (118)

exactly as being claimed.

According to (118), the l.h.s. of (110) and (115) are equal. However, the r.h.s. of (110) differs from the r.h.s. of (115), the latter being $\rho$-dependent. This means in turn that statement (33) of our basic Theorem II does generally depend upon the coordinates chosen. Coordinate transformations can be thus exploited to optimize (strengthen) inequality (33).

6. Conclusion and Prospects

In summary, our basic Theorem II stated above generalizes a recently derived random inequality of [20] to the case of multidimensional random variables. Six subsequent additional results (Theorems I-1–I-2, Theorems II-1–II-4) apply then Theorem II to different frequently encountered statistical distributions (multiuniform, multinomial, normal, and multinormal). Furthermore, we show in Section 5 that basic inequality (33) of Theorem II is coordinate dependent (and thus optimizable via carrying out suitable coordinate transformations). The just mentioned formulas and insights could be useful for making estimates in multivariate statistics.

Finally, we find it useful to list below three open questions which may be worthy of examining in the future. Namely, (i) we have assumed above that $\Omega$ is a convex open set. Any nonconvex path connected open set $\widetilde{\Omega}$ can actually be made convex via a suitable coordinate transformation. So Theorem II turns out to be even more general than stated in Section 3. It is not a priori clear, however, what would happen in the case when $\Omega$ is a disconnected open set. This remains to be seen.
(ii) an optimization of inequality (33) via coordinate transformations represents a very promising direction of further research

(iii) it might be desirable to supplement some real life application which would reflect the practical value of our inequality (33). For example, one may think about applications in physics or economy

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Hainan Province (Grant no. 2018CXTD338; contribution rate: 30%), the National Natural Science Foundation of China (Grant no. 11761027; contribution rate: 50%), the Scientific Research Foundation of Hainan Province Education Bureau (Grant no. Hnky2016-14; contribution rate: 10%), and the Educational Reform Foundation of Hainan Province Education Bureau (Grant no. Hnjg2017ZD-13; contribution rate: 10%).

References


