

Relationship Between Stochastic Inequalities and Some Classical Mathematical Inequalities*

Y.L. TONG

*School of Mathematics, Georgia Institute of Technology,
Atlanta, GA 30332-0160, USA
e-mail: tong@math.gatech.edu*

(Received 2 May 1996)

The notions of association and dependence of random variables, rearrangements, and heterogeneity via majorization ordering have proven to be most useful for deriving stochastic inequalities. In this survey article we first show that these notions are closely related to three basic inequalities in classical mathematical analysis: Chebyshev's inequality, the Hardy-Littlewood-Pólya rearrangement inequality and Schur functions. We then provide a brief review of some of the recent results in this area. An overall objective is to illustrate that classical mathematical inequalities of this type play a central role in the developments of stochastic inequalities.

Keywords: stochastic inequalities; association; rearrangements; majorization and Schur functions.

1991 AMS Subject Classifications: Primary 60E15, Secondary 62H99.

1 INTRODUCTION

As noted by Pólya [21], "Inequalities play a role in most branches of mathematics and have different applications". This is certainly true in the area of probability and statistics. For example, the celebrated Chebyshev, Markov and Kolmogorov inequalities are well-known and can be found in many probability books. In the theory of estimation, the Cramér-Rao inequality and its generalizations provide lower bounds on the variances of a large class

*Research supported in part by NSA Grant No. MDA904-94-H-2032.

of estimators. In statistical hypotheses testing, the Neyman-Pearson Lemma directly involves inequalities.

It appears that the developments of many of the stochastic inequalities originated from the concepts of certain fundamental inequalities in mathematics. As a result, mathematical inequalities of this type have made a strong impact on stochastic inequalities. In this survey paper, we discuss the influence of three basic inequalities in classical mathematical analysis, and show how certain stochastic inequalities are closely related to those classical results. Specifically, we explain how and why the concepts of (1) Chebyshev's (other) inequality, (2) the Hardy-Littlewood-Pólya (HLP) rearrangement inequality and (3) the Schur functions are related to, respectively, (1') inequalities via association and dependence of random variables, (2') stochastic inequalities via arrangement increasing functions, and (3') stochastic inequalities via majorization. Those results are described and discussed in Sections 2, 3 and 4, respectively. For reader's convenience, each of the three classical results is restated at the beginning of a section. Some related stochastic inequalities, which have been obtained during the past two decades, are then stated with selected applications in Section 5. For simplicity we assume that all functions and subsets involved are Borel-measurable and integrable. Nondecreasing (nonincreasing) functions will be called increasing (decreasing). Further, we note that the results described in this paper are for illustrative purposes only, hence they are neither complete nor exclusive. For additional results on these topics, the reader is referred to the bibliographies contained in the references listed at the end of this paper. The proofs of some of the results can also be found there.

2 CHEBYSHEV'S INEQUALITY, ASSOCIATION, AND STOCHASTIC DEPENDENCE

We first state a classical result of Chebyshev; a convenient reference for this result is Hardy *et al.* ([9], p. 43). For notational consistency, the statement given here is in a modified version and is not in its original form.

THEOREM 2.1 *Let $G_1, G_2 : [0, 1] \rightarrow \mathbb{R}$ be two real-valued functions. If they are both increasing, then the inequality*

$$\int_0^1 G_1(u)G_2(u)du \geq \left[\int_0^1 G_1(u)du \right] \left[\int_0^1 G_2(u)du \right] \quad (2.1)$$

holds.

Note that in Theorem 2.1 if the word “increasing” is replaced by “decreasing” then, by replacing G_j with $-G_j$ ($j = 1, 2$) in (2.1), the same conclusion also holds. When compared with his inequality that provides bounds for the tail probabilities of a distribution via the first two moments, this result seems to be less well-known. Thus Theorem 2.1 is sometimes referred to as “Chebyshev’s other inequality” in the literature. If we interpret the integrals as expectations of random variables, then this result yields a corresponding stochastic inequality in the following sense: Consider a random variable X with distribution function $F(x)$, and let $(g_1(X), g_2(X))$ be a two-dimensional random vector such that g_1, g_2 are real-valued functions. By letting $u = F(x)$ and $G_j = g_j F^{-1}$ ($j = 1, 2$) in (2.1), we immediately obtain

THEOREM 2.1’. *If g_1, g_2 are both increasing or both decreasing, then*

$$E[g_1(X)g_2(X)] \geq [E(g_1(X))][E(g_2(X))] \quad (2.1')$$

holds; or equivalently,

$$\text{Corr}(g_1(X), g_2(X)) \geq 0, \quad (2.1'')$$

where $E(\cdot)$ denotes expectation and “Corr” stands for the correlation coefficient.

Intuitively speaking, Theorem 2.1’ states that if g_1 and g_2 are both increasing or both decreasing, then $g_1(X)$ and $g_2(X)$ tend to take larger values together and smaller values together. Thus their correlation coefficient is nonnegative. The theorem then involves positive dependence of random variables which are monotone functions of a common random variable. A question of interest is whether it can be generalized to monotone functions of *several* random variables. This question can be answered by studying the following concept of association of random variables, which was first considered by Esary, *et al.* [8]:

DEFINITION 2.2. For $n \geq 1$ the random variables X_1, \dots, X_n are said to be associated, or the set of random variables $\{X_1, \dots, X_n\}$ is said to be associated, if for all given real-valued functions g_1, g_2 that are increasing in each component when the other components are held fixed, the inequality

$$E \left[\prod_{j=1}^2 g_j(\mathbf{X}) \right] \geq \prod_{j=1}^2 E(g_j(\mathbf{X}))$$

holds, or equivalently,

$$\text{Corr}(g_1(\mathbf{X}), g_2(\mathbf{X})) \geq 0, \quad (2.2)$$

where $\mathbf{X} = (X_1, \dots, X_n)$.

Esary, *et al.* [8] proved the following theorem:

THEOREM 2.3. (a) A set consisting of a single random variable is a set of associated random variables. (b) Independent random variables are associated random variables. (c) A subset of a set of associated random variables forms a set of associated random variables. (d) Increasing functions of associated random variables are associated random variables.

From a historical viewpoint, the Esary-Proschan-Walkup paper was motivated by a research problem in application and was not a direct outgrowth of the Chebyshev inequality. But, nevertheless, the notion is clearly a generalization of that in Theorem 2.1'. In fact, Theorem 2.3(a) is equivalent to Theorem 2.1'.

A repeated application of Theorem 2.3 yields the following multivariate probability inequality. Note that, in particular, the result applies to independent random variables.

FACT 2.4. For $k \geq 2$ let $g_1, \dots, g_k : \mathbb{R}^n \rightarrow \mathbb{R}$ be increasing. If X_1, \dots, X_n are associated random variables, then

$$\begin{aligned} P \left[\bigcap_{j=1}^k \{g_j(\mathbf{X}) \geq \lambda_j\} \right] &\geq P \left[\bigcap_{j=1}^{k'} \{g_j(\mathbf{X}) \geq \lambda_j\} \right] P \left[\bigcap_{j=k'+1}^k \{g_j(\mathbf{X}) \geq \lambda_j\} \right] \\ &\geq \prod_{j=1}^k P [g_j(\mathbf{X}) \geq \lambda_j] \end{aligned} \quad (2.3)$$

holds for all $k' < k$ and arbitrary but fixed real numbers $\lambda_1, \dots, \lambda_k$.

Fact 2.4 has a special application to stochastic processes with independent increments:

FACT 2.5. Let $\mathcal{S} = \{Y_t : t \in T\}$ be a stochastic process of independent increments, where the parameter space T is either discrete or continuous. For $k \geq 2$ let $g_1, \dots, g_k : \mathbb{R}^n \rightarrow \mathbb{R}$ be increasing functions. Then for arbitrary but fixed $t_1 < \dots < t_n$ in T the set of random variables $\{g_1(Y_{t_1}, \dots, Y_{t_n}), \dots, g_k(Y_{t_1}, \dots, Y_{t_n})\}$ is associated. Thus, when substituting $(Y_{t_1}, \dots, Y_{t_n})$ for \mathbf{X} in (2.3), the inequalities hold for all $k' < k$ and all $\lambda_1, \dots, \lambda_k$.

Fact 2.5 has certain applications in boundary-crossing type of problems, and it represents a generalization of some earlier results. For example, Robbins [23] previously obtained the last inequality in (2.3) for $k = n$ and $g_j(Y_{t_1}, \dots, Y_{t_n}) = Y_{t_j}$ ($j = 1, \dots, k$) via a direct verification. His result now follows from Fact 2.5 as a special case.

The notion of association and monotone transformations of random variables arise in many problems in probability and statistics. For an extensive application in reliability theory, especially on reliability bounds for coherent systems, see Barlow and Proschan [1, Chapter 2]. In multivariate statistical analysis, Fact 2.4 is often applied to reduce the dimensionality of the joint distribution of random variables. In this area there are other related notions for defining positive dependence of random variables. A notion that is stronger than association is the multivariate totally-positive-of-order-2 (MTP₂) property of the joint probability density function (p.d.f.). For a comprehensive treatment of TP₂ functions and related results, see Karlin [14]. For a description of the orderings of the notions of positive dependence, see Barlow and Proschan [1, Chapter 5] and Tong [28, Chapter 5]. It is known that for the multivariate normal distribution all of these notions are equivalent. As a result, many special results for this distribution have been obtained. For example, Pitt's [20] result states that if $\mathbf{X} = (X_1, \dots, X_n)$ is distributed according to a multivariate normal distribution, then X_1, \dots, X_n are associated if and only if all of the simple correlation coefficients are nonnegative.

3 THE HLP REARRANGEMENT INEQUALITY AND ARRANGEMENT INCREASING (AI) FUNCTIONS

The HLP rearrangement inequality, due to Hardy *et al.* [9, Chapter 10], is an algebraic inequality for inner products. For $n \geq 2$ let $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{b} = (b_1, \dots, b_n)$ be two real vectors. Their inner product is defined to be $\mathbf{a}\mathbf{b}^T = \sum_{i=1}^n a_i b_i$. The problem of interest is to find the maximum and minimum through all permutations of the components of \mathbf{a} and \mathbf{b} . The HLP rearrangement inequality states that

THEOREM 3.1. *Let $(a_{[1]}, \dots, a_{[n]})$ and $(a_{(1)}, \dots, a_{(n)})$ be two vectors obtained from rearranging the components of \mathbf{a} such that $a_{[1]} \geq \dots \geq a_{[n]}$ and $a_{(1)} \leq \dots \leq a_{(n)}$, and let $(b_{[1]}, \dots, b_{[n]})$, $(b_{(1)}, \dots, b_{(n)})$ be defined similarly. Then for all \mathbf{a} , \mathbf{b} the inequalities*

$$\sum_{i=1}^n a_{(i)} b_{[i]} \leq \sum_{i=1}^n a_i b_i \leq \sum_{i=1}^n a_{(i)} b_{(i)}$$

hold. Thus, the inner product of \mathbf{a} , \mathbf{b} is minimized (maximized) over all possible permutations when their components are completely reversely ordered (similarly ordered).

Without loss of generality, let us assume that $a_1 \leq a_2 \leq \dots \leq a_n$ already holds. We shall denote such a vector by the symbol $\mathbf{a} \uparrow$. If the components of \mathbf{b} are not in an ascending order, then there exist two integers $r < s$ such that $b_r > b_s$. Let \mathbf{b}_1 denote the vector obtained from \mathbf{b} by interchanging b_r and b_s in \mathbf{b} . We use the symbol $\mathbf{b} \stackrel{p}{<} \mathbf{b}_1$ to mean that b_r and b_s are now in an ascending order in the new vector \mathbf{b}_1 . If $\mathbf{b}^* = (b_1^*, \dots, b_n^*)$ is another vector such that there exists N vectors $\mathbf{b}_1, \dots, \mathbf{b}_N$ satisfying $\mathbf{b} \stackrel{p}{<} \mathbf{b}_1 \stackrel{p}{<} \dots \stackrel{p}{<} \mathbf{b}_N \stackrel{p}{<} \mathbf{b}^*$, then we can change \mathbf{b} into \mathbf{b}^* , in a finite number of rearrangements by interchanging two components at a time in this fashion. In this case we write $\mathbf{b} \stackrel{pp}{<} \mathbf{b}^*$. Adopting such an approach, Sobel [26] obtained the following generalization of Theorem 3.1:

THEOREM 3.1'. *Let \mathbf{a} , \mathbf{b} and \mathbf{b}^* be n -dimensional real vectors. If $\mathbf{b} \stackrel{pp}{<} \mathbf{b}^*$, then $(\mathbf{a} \uparrow) \mathbf{b}^T \leq (\mathbf{a} \uparrow) (\mathbf{b}^*)^T$.*

In a 1977 paper, Hollander, Proschan and Sethuraman studied functions that are decreasing in transposition. Subsequently, such functions are called "arrangement increasing" in the Marshall-Olkin [17, Section 6.F) book. The definition of such a function is given below:

DEFINITION 3.2. $\phi(\mathbf{a}, \mathbf{b}) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be an arrangement increasing (AI) function of (\mathbf{a}, \mathbf{b}) if (i) $\phi(\mathbf{a}, \mathbf{b}) = \phi(\boldsymbol{\pi}(\mathbf{a}), \boldsymbol{\pi}(\mathbf{b}))$ for every vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ which is a permutation of $(1, \dots, n)$, where $\boldsymbol{\pi}(a_1, \dots, a_n) = (a_{\pi_1}, \dots, a_{\pi_n})$; (ii) $\mathbf{b} \stackrel{pp}{<} \mathbf{b}^*$ implies $\phi(\mathbf{a} \uparrow, \mathbf{b}) \leq \phi(\mathbf{a} \uparrow, \mathbf{b}^*)$.

With this definition, Theorem 3.1' essentially says that $\phi(\mathbf{a}, \mathbf{b}) = \mathbf{a} \mathbf{b}^T$ is an AI function. When applied to probability and statistics, the problem of interest is to develop some basic results for showing that certain functions are AI, thus to obtain probability inequalities via rearrangements. Hollander *et al.* [10] proved a fundamental preservation theorem and gave many such results. In particular, they proved a theorem that involves the ranks and the parameters

of random variables from several populations. Let $\mathcal{F} = \{f(x, \theta) : \theta \in \Omega\}$ denote a family of univariate p.d.f.'s while $\Omega \subset \mathbb{R}$ is the parameter space. \mathcal{F} is said to possess the monotone likelihood ratio (MLR) property if, for all $\theta_1 < \theta_2$ in Ω the ratio $f(x, \theta_2)/f(x, \theta_1)$ is an increasing function of x . This property represents a notion of stochastic ordering of random variables in the sense that $E_\theta g(X)$ is increasing in θ for all increasing functions g . Thus a random variable X associated with a larger θ is stochastically larger in this sense. Now, let X_1, \dots, X_n be n independent random variables with p.d.f.'s $f(x, \theta_1), \dots, f(x, \theta_n)$, respectively. For $i = 1, \dots, n$ let R_i denote the rank of X_i in the combined sample (X_1, \dots, X_n) . Then, intuitively speaking, the rank of a random variable with a larger θ value tends to be higher. This is affirmed in the following theorem due to Hollander *et al.* [10]:

THEOREM 3.3. *If \mathcal{F} has the MLR property and X_1, \dots, X_n are independent, then*

$$\phi(\boldsymbol{\theta}, \mathbf{r}) = P_{\boldsymbol{\theta}}[R_1 = r_1, \dots, R_n = r_n]$$

is an AI function of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ and $\mathbf{r} = (r_1, \dots, r_n)$.

Theorem 3.3 has many useful applications in rank order statistics and nonparametrics. For details, see Hollander *et al.* [10] and the related references. Other stochastic applications of arrangement functions include statistical inference problems based on order statistics. In particular, the ranking and selection problems concern the selection of the populations associated with the larger parameters. By applying Theorem 3.1', it can be shown algebraically that if the family of p.d.f.'s, \mathcal{F} , is an exponential family, then the corresponding likelihood function is an AI function of $(\boldsymbol{\theta}, \mathbf{x})$; thus it is maximized when the parameters $\theta_1, \dots, \theta_n$ and the observations x_1, \dots, x_n are similarly ordered. This property was applied extensively by Bechhofer, Kiefer and Sobel in their 1968 monograph.

4 MAJORIZATION, SCHUR FUNCTIONS, AND STOCHASTIC INEQUALITIES

The notion of majorization concerns the heterogeneity of the elements of a real vector. Let $\mathbf{a} = (a_1, \dots, a_n)$, $\mathbf{b} = (b_1, \dots, b_n)$ be two real vectors and the $a_{[i]}$'s, $b_{[i]}$'s be defined as in Theorem 3.1.

DEFINITION 4.1. \mathbf{a} is said to majorize \mathbf{b} , in symbols $\mathbf{a} \succ \mathbf{b}$, if $\sum_{i=1}^m a_{[i]} \geq \sum_{i=1}^m b_{[i]}$ holds for $m = 1, \dots, n - 1$ and equality holds for $m = n$.

This definition provides a partial ordering of the “heterogeneity” of the components of \mathbf{a} and \mathbf{b} . In particular, it is easy to see that (i) $\mathbf{a} \succ (\bar{a}, \dots, \bar{a})$ holds for all \bar{a} where \bar{a} is the arithmetic mean, (ii) if $a_i \geq 0$ ($i = 1, \dots, n$), then $(\sum_{i=1}^n a_i, 0, \dots, 0) \succ \mathbf{a}$. It is known that:

THEOREM 4.2. (a) $\mathbf{a} \succ \mathbf{b}$ holds iff $\prod_{i=1}^n h(a_i) \geq \prod_{i=1}^n h(b_i)$ holds for all continuous convex functions h . (b) $\mathbf{a} \succ \mathbf{b}$ holds iff $\mathbf{b} = \mathbf{a}P$ for some $n \times n$ doubly stochastic matrix P .

Definition 4.1 is closely related to the following definition of Schur-convex and Schur-concave functions:

DEFINITION 4.3. $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be a Schur-convex (Schur-concave) function if $\mathbf{a} \succ \mathbf{b}$ implies $\psi(\mathbf{a}) \geq \psi(\mathbf{b})$ ($\psi(\mathbf{a}) \leq \psi(\mathbf{b})$).

The basic ideas of majorization and Schur functions have played a role in classical mathematical analysis. In their monograph, Marshall and Olkin [17] gave a comprehensive and complete treatment of this topic, including its historical developments, basic theory and related inequalities and various applications that were known before 1979.

In the rest of this section we provide a review of some existing stochastic inequalities derived via majorization. An earlier result, due to Marshall and Olkin [16], can be stated in the form of an integral inequality:

THEOREM 4.4. If f and g are Schur-concave functions defined on \mathbb{R}^n , then the function defined on \mathbb{R}^n by

$$\psi(\boldsymbol{\theta}) = \int_{\mathbb{R}^n} g(\boldsymbol{\theta} - \mathbf{x}) f(\mathbf{x}) d\mathbf{x}$$

is a Schur-concave function.

By using the property that $(\boldsymbol{\theta} - \mathbf{x}) \succ (\boldsymbol{\theta}^* - \mathbf{x}^*)$ iff $(\mathbf{x} - \boldsymbol{\theta}) \succ (\mathbf{x}^* - \boldsymbol{\theta}^*)$, and by letting g be the indicator function of a Schur-concave subset in \mathbb{R}^n , Theorem 4.4 implies the following result given in Marshall and Olkin [16]:

FACT 4.5. If the p.d.f. $f(\mathbf{x})$ of an n -dimensional random vector $\mathbf{X} = (X_1, \dots, X_n)$ is Schur-concave, then the distribution function $F(\mathbf{a}) = P[X_1 \leq a_1, \dots, X_n \leq a_n]$ is a Schur-concave function of $\mathbf{a} = (a_1, \dots, a_n)$.

Proschan and Sethuraman [22] applied the results in majorization theory to study stochastic majorization and provided the following definition:

DEFINITION 4.6. A random vector $\mathbf{X} = (X_1, \dots, X_n)$ is said to stochastically majorize $\mathbf{Y} = (Y_1, \dots, Y_n)$ if $E(\psi(\mathbf{X})) \geq (\leq) E(\psi(\mathbf{Y}))$ holds for all Schur-convex (Schur-concave) functions ψ .

They then obtained several useful theorems via this stochastic majorization ordering. One such result deals with the probability contents of a Schur-concave set for nonnegative random variables with TP_2 and semigroup properties. A convenient reference of that theorem and its applications is Marshall and Olkin [17, p. 101). Hollander *et al.* [10] also showed that if $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is of the form $\phi(\mathbf{a}, \mathbf{b}) = g(\mathbf{a} - \mathbf{b})$ for some function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, then ϕ is AI iff $g(\mathbf{x})$ is a Schur-concave function of \mathbf{x} . This result yields rearrangement inequalities via majorization for location parameter families in statistics.

5 SOME RECENT DEVELOPMENTS

There are many recent results in stochastic inequalities that have been derived through association of random variables, AI functions and majorization theory. Due to space limitations, we describe only a few of them in this section. Interested readers may find additional results from the bibliographies contained in the references given in this paper.

5.1 Positive Association and Negative Association

The random variables that satisfy Definition 2.2 tend to take larger values together and smaller values together. Thus they may be called “positively associated”. Many useful inequalities derived via this notion have become available after the publication of the Esary *et al.* [8] paper, and some of them are originated from applications. A comprehensive reference on their applications in reliability theory can be found in the Barlow-Proschan [1] book. More recent applications in this area can be found in the review article by Boland *et al.* [6]. Some of the recent results and their selected applications in biology and medicine, business and economics, operations research (including queueing theory), and statistical inference can be found in Block *et al.* [3], Shaked and Tong [25] and Shaked and Shanthikumar [24].

The concept of negative association was introduced by Block *et al.* [4], Joag-Dev and Proschan [11] and others. There have been different forms of the definition of negative association and a convenient one is the following (Joag-Dev and Proschan [11]):

DEFINITION 5.1. Random variables X_1, \dots, X_n are said to be negatively associated if for all disjoint subsets $C_1 = \{r_1, \dots, r_m\}$ and $C_2 = \{s_1, \dots, s_{m'}\}$ of $\{1, \dots, n\}$ and all increasing functions $g_1 : \mathbb{R}^m \rightarrow \mathbb{R}$ and $g_2 : \mathbb{R}^{m'} \rightarrow \mathbb{R}$ the inequality

$$\begin{aligned} E [g_1 (X_{r_1}, \dots, X_{r_m}) g_2 (X_{s_1}, \dots, X_{s_{m'}})] &\leq \\ E [g_1 (X_{r_1}, \dots, X_{r_m})] E [g_2 (X_{s_1}, \dots, X_{s_{m'}})] & \end{aligned}$$

holds.

It follows immediately from Definition 5.1 that if X_1, \dots, X_n are negatively associated, then

$$\begin{aligned} P \left[\bigcap_{i=1}^m \{X_{r_i} \geq \lambda_i\}, \bigcap_{i=1}^{m'} \{X_{s_i} \geq \delta_i\} \right] &\leq \\ P \left[\bigcap_{i=1}^m \{X_{r_i} \geq \lambda_i\} \right] P \left[\bigcap_{i=1}^{m'} \{X_{s_i} \geq \delta_i\} \right] & \end{aligned}$$

for all λ_i and δ_i . Certain applications of negative association of random variables to specific distributions have been studied in the literature. For example, Block *et al.* [4] proved that if the joint distribution of X_1, \dots, X_n is multinomial, then they are negatively associated.

5.2 Stochastic Arrangement Increasing and Multivariate Arrangement Increasing Functions

Certain generalizations of the notion of arrangement increasing functions have been made recently for deriving new multivariate probability inequalities. For example, Boland *et al.* [6] applied the fundamental preservation theorem in Hollander *et al.* [10] to obtain the following result:

THEOREM 5.2. Assume that (X_1, \dots, X_n) is a random vector with a p.d.f. that is permutation invariant. Let h_1, h_2 be AI functions on $\mathbb{R}^n \times \mathbb{R}^n$, and let $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ be increasing. Then

$$\phi(\mathbf{a}, \mathbf{b}) = E[g_1(h_1(\mathbf{a}, \mathbf{X}))g_2(h_2(\mathbf{X}, \mathbf{b}))]$$

is an AI function of $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^n \times \mathbb{R}^n$, where the expectation is taken over the distribution of \mathbf{X} .

By applying this result to permutation invariant distributions and to families of probability distributions with AI density functions, many new results were given in Boland *et al.* [6]. For example, Theorem 5.2 immediately implies that if $f(\mathbf{x})$ is permutation invariant, then

$$P[a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n]$$

is an AI function of (\mathbf{a}, \mathbf{b}) . (This is an earlier unpublished result of Boland.) For a convenient reference of these results, see Chapter 16 of Pečarić *et al.* [19].

Another set of new results on this topic were given by Boland and Proschan [5] through multivariate generalization of AI functions. They provided a useful definition of multivariate AI functions of the form $\phi : \mathbb{R}^n \times \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}$, proved a main theorem, showed that many existing classical results are special cases of the main theorem, and then applied it to derive new results in multivariate probability inequalities. For details, see Boland and Proschan [5] or Chapter 17 of Pečarić *et al.* [19].

5.3 Some Recent Results and Generalizations of Majorization-Related Inequalities

In the area of majorization-related stochastic inequalities, many new results have become available since the publication of the Marshall-Olkin [17] book. The following is a description of some of them:

5.3 (A) Generalized Majorization and Multivariate Majorization

There exist several generalizations of the notion of majorization, including multivariate majorization, majorization orderings for continuous functions, and other related orderings. Three types of multivariate majorization were already treated in Chapter 15 of Marshall and Olkin [17], and useful stochastic inequalities have been obtained via their applications. A generalization of majorization ordering via integrals of continuous functions and other multivariate majorization ideas were treated recently by Joe [12], Joe and Verducci [13], and others. Selected applications to the comparison of stochastic heterogeneity of probability distributions were also given.

5.3 (B) Additional Majorization-Related Multivariate Probability Inequalities

Many additional multivariate inequalities have been obtained via majorization ordering. For example, motivated by Fact 4.5 Tong [29] proved that:

FACT 5.3. *If the p.d.f. $f(\mathbf{x})$ of $\mathbf{X} = (X_1, \dots, X_n)$ is Schur-concave, then (a) the distribution function of $|\mathbf{X}| = (|X_1|, \dots, |X_n|)$, $P[|X_1| \leq a_1, \dots, |X_n| \leq a_n]$, is a Schur-concave function of $\mathbf{a} = (a_1, \dots, a_n)$; (b) the probability $P[\sum_{i=1}^n (X_i/a_i)^2 \leq \lambda]$ is a Schur-concave-function of (a_1^2, \dots, a_n^2) for every fixed $\lambda > 0$.*

Tong [29] also conjectured that if the conditions on $f(\mathbf{x})$ in Fact 5.3 are satisfied, then for all positive even integers $m \geq 2$ the probability $P[\sum_{i=1}^n (X_i/a_i)^m \leq \lambda]$ is a Schur-concave function of $(a_1^{m/(m-1)}, \dots, a_n^{m/(m-1)})$ for all $\lambda > 0$. In a 1983 paper Karlin and Rinott proved that this conjecture is true. Karlin and Rinott [15] and Tong [30] also independently obtained a multivariate probability inequality for n -dimensional rectangles via multivariate majorization. For a description of that and other related results, see the survey articles Tong [31] and Tong [33]. We note that all of these results yield integral inequalities in \mathbb{R}^n .

5.3 (C) A Positive Dependence Ordering via Majorization Ordering of Dimension Vectors

In a 1977 paper, Tong applied Muirhead's inequality (see Marshall and Olkin [17, p. 87]) to prove the following result:

THEOREM 5.4. *Let X_1, \dots, X_n be nonnegative random variables. If their joint p.d.f. is absolutely continuous with respect to either Lebesgue measure or the counting measure, and is permutation invariant, then $E \prod_{i=1}^n (X_i)^{a_i}$ is a Schur-convex function of $\mathbf{a} = (a_1, \dots, a_n)$ ($EX^0 \equiv 1$).*

A simple application of Theorem 5.4 yields the following moment inequality:

FACT 5.5. *If X is a nonnegative random variable, either continuous or discrete, then $\prod_{i=1}^n \mu_{a_i}$ is a Schur-convex function of \mathbf{a} , where $\mu_{a_i} = EX^{a_i}$ and $\mu_0 \equiv 1$.*

Tong [32] applied these results to a large class of distributions for obtaining a partial ordering of positive dependence of random variables via a majorization ordering of the dimension vectors. Detailed applications to certain families of probability distributions and stochastic processes were given. In a recent paper, Olkin and Tong [18] applied these results to study the effects of positive dependence in reliability theory and shock models when the random variables are exponentially distributed.

References

- [1] R.E. Barlow and F. Proschan (1975, 1981), *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York and To Begin With, Silver Spring, MD.
- [2] R.E. Bechhofer, J. Kiefer and M. Sobel (1968), *Sequential Identification and Ranking Procedures*. University of Chicago Press, Chicago, IL.
- [3] H.W. Block, A.R. Sampson and T.H. Savits, eds. (1990), *Topics in Statistical Dependence*. The Institute of Mathematical Statistics, Hayward, CA.
- [4] H.W. Block, T.H. Savits and M. Shaked (1982), Some concepts of negative dependence. *Ann. Probab.*, **10**, 765–772.
- [5] P.J. Boland and F. Proschan (1988), Multivariate arrangement increasing functions with applications in probability and statistics. *J. Multivariate Anal.*, **25**, 286–298.
- [6] P.J. Boland, F. Proschan and Y.L. Tong (1988), Moment and geometric probability inequalities arising from arrangement increasing functions. *Ann. Probab.*, **16**, 407–413.
- [7] P.J. Boland, F. Proschan and Y.L. Tong (1993), Some recent applications of stochastic inequalities in system reliability theory. In *Advances in Reliability*, A.P. Basu, ed. Elsevier Science Publishers, New York, pp. 29–41.
- [8] J.D. Esary, F. Proschan and D.W. Walkup (1967), Association of random variables, with applications. *Ann. Math. Statist.*, **38**, 1466–1474.
- [9] G.H. Hardy, J.E. Littlewood and G. Pólya (1934, 1952), *Inequalities*. 1st ed., 2nd ed. Cambridge University Press, Cambridge, England.
- [10] M. Hollander, F. Proschan and J. Sethuraman (1977), Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist.*, **5**, 722–733.
- [11] K. Joag-Dev and F. Proschan (1983), Negative association of random variables, with applications. *Ann. Statist.*, **11**, 286–295.
- [12] H. Joe (1993), Generalized majorization orderings and applications. In *Stochastic Inequalities*, M. Shaked and Y. L. Tong, eds. Institute of Mathematical Statistics, Hayward, CA, pp. 145–158.
- [13] H. Joe and J.V. Verducci (1993), Multivariate majorization by positive combinations. In *Stochastic Inequalities*, M. Shaked and Y. L. Tong, eds. Institute of Mathematical Statistics, Hayward, CA, pp. 159–181.
- [14] S. Karlin (1968), *Total Positivity*, Vol. 1. Stanford University Press, Stanford, CA.
- [15] S. Karlin and Y. Rinott (1983), Comparison of measures, multivariate majorization, and applications to statistics. In *Studies in Econometrics, Time Series and Multivariate Analysis*, S. Karlin, T. Amemiya and L.A. Goodman, eds. Academic Press, New York, pp. 465–489.
- [16] A.W. Marshall and I. Olkin (1974), Majorization in multivariate distributions. *Ann. Statist.*, **2**, 1189–1200.
- [17] A.W. Marshall and I. Olkin (1979), *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

- [18] I. Olkin and Y.L. Tong (1994), Positive dependence of a class of multivariate exponential distributions. *SIAM J. Cont. Opt.*, **32**, 965–974.
- [19] J.E. Pečarić, F. Proschan and Y.L. Tong (1992), *Convex Functions, Partial Orderings, and Statistical Applications*. Academic Press, Boston.
- [20] L.D. Pitt (1982), Positively correlated normal variables are associated. *Ann. Probab.*, **10**, 496–499.
- [21] G. Pólya (1967), Inequalities and the principle of nonsufficient reason. In *Inequalities*, O. Shisha, ed. Academic Press, New York, 1–15.
- [22] F. Proschan and J. Sethuraman (1977), Schur functions in statistics. I. The Preservation Theorem. *Ann. Statist.*, **5**, 256–262.
- [23] H. Robbins (1954), A remark on the joint distribution of cumulative sums. *Ann. Math. Statist.*, **25**, 614–616.
- [24] M. Shaked and J.G. Shanthikumar (1994), *Stochastic Orders and Their Applications*. Academic Press, Boston.
- [25] M. Shaked and Y.L. Tong, eds. (1993), *Stochastic Inequalities*. Institute of Mathematical Statistics, Hayward, CA.
- [26] M. Sobel (1954), On a generalization of an inequality of Hardy, Littlewood and Pólya. *Proc. Amer. Math. Soc.*, **5**, 596–602.
- [27] Y.L. Tong (1977), An ordering theorem for conditionally independent and identically distributed random variables. *Ann. Statist.*, **5**, 274–277.
- [28] Y.L. Tong (1980), *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.
- [29] Y.L. Tong (1982), Rectangular and elliptical probability inequalities for Schur-concave random variables. *Ann. Statist.*, **10**, 637–642.
- [30] Y.L. Tong (1983, 1989), Probability inequalities for n -dimensional rectangles via multivariate majorization. *Tech. Report No. 189*, (1983), Department of Statistics, Stanford University, Stanford, CA. Revised version in *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin*. L.J. Gleser, M.D. Perlman, S.J. Press and A.R. Sampson, eds. Springer-Verlag (1989), New York and Berlin, pp. 146–159.
- [31] Y.L. Tong (1988), Some majorization inequalities in multivariate statistical analysis. *SIAM Review*, **30**, 602–622.
- [32] Y.L. Tong (1989), Inequalities for a class of positively dependent random variables with a common marginal. *Ann. Statist.*, **17**, 429–435.
- [33] Y.L. Tong (1994), Some recent developments on majorization inequalities in probability and statistics. *Linear Alg. Appl.*, **199**, 69–90.