The Annals of Statistics
1977, Vol. 5, No. 2, 263–273

SCHUR FUNCTIONS IN STATISTICS II. STOCHASTIC MAJORIZATION[1]

BY S. E. NEVIUS, F. PROSCHAN AND J. SETHURAMAN

Florida State University

This is Part II of a two-part paper. The main purpose of this two-part paper is (a) to develop new concepts and techniques in the theory of majorization and Schur functions, and (b) to obtain fruitful applications in probability and statistics. In Part II we introduce a stochastic version of majorization, develop its properties, and obtain multivariate applications of both the preservation theorem of Part I and the new notion of stochastic majorization. This leads to a definition of Schur families of multivariate distributions. Generalizations are obtained of earlier results of Olkin and of Wong and Yue; in addition, new results are obtained for the multinomial, multivariate negative binomial, multivariate hypergeometric, Dirichlet, negative multivariate hypergeometric, and multivariate logarithmic series distributions.

1. Introduction and summary. In Part I we derived a basic theorem concerning the preservation of a Schur function under certain integral transformations. (Definitions, notation, and conventions of Part II are as in Part I and will generally not be repeated.) In Part II we introduce a stochastic version of majorization, develop its properties, and obtain multivariate applications of both the preservation theorem of Part I and the new notion of stochastic majorization.

2. Stochastic majorization: definition and characterizations. Throughout the paper, let $\mathbf{X}$ and $\mathbf{X}'$ be random vectors taking values in R_n, and let P and P' be the probability measures on the Borel subsets of R_n generated by $\mathbf{X}$ and $\mathbf{X}'$ respectively.

DEFINITIONS.

(1) A random vector $\mathbf{X}$ *stochastically majorizes* a random vector $\mathbf{X}'$ if $f(\mathbf{X}) \geqq^{\text{st}} f(\mathbf{X}')$ for every Borel measurable Schur-convex function f on R_n; in symbols, $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$.

(2) Probability measure P *stochastically majorizes* probability measure P' if $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$; in symbols $P \geqq^{\text{st.m.}} P'$.

Notice that Definition (1) is a stochastic analogue of a characterization of

Received August 1974; revised July 1976.

[1] The research of Professors Nevius and Proschan was sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-74-2581A. The research of Professor Sethuraman was supported by the United States Army Research Office, Durham, under Grant No. DAHC04 74 G 0188. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

AMS 1970 *subject classifications.* 26A51, 26A86, 60E05, 62H99.

Key words and phrases. Majorization, Schur-concave, Schur-convex, Schur functions, stochastic majorization, multivariate distributions, inequalities.

deterministic majorization, namely that $\mathbf{x} \geqq^{\mathrm{m}} \mathbf{x}'$ if and only if $f(\mathbf{x}) \geqq f(\mathbf{x}')$ for every Schur-convex function f.

Although several stochastic analogues of other versions of the definition of majorization in the deterministic case are possible, we will see later on in this section that they are not all equivalent. The definition given above will be seen to have advantages over the alternative definitions.

This section is devoted to the characterization of stochastic majorization and the study of some of its consequences.

LEMMA 2.1. *Let* $\mathbf{X} \geqq^{\mathrm{st.m.}} \mathbf{X}'$. *Let* $S = \sum_1^n X_i$ *and* $S' = \sum_1^n X_1'$ *Then* $S =^{\mathrm{st}} S'$.

PROOF. Notice that both $s(\mathbf{x}) = \sum_1^n x_i$ and $-s(\mathbf{x})$ are Schur-convex. Thus $S \geqq^{\mathrm{st}} S'$ and $S \leqq^{\mathrm{st}} S'$. Hence $S =^{\mathrm{st}} S'$. □

DEFINITION. A subset A of R_n is said to be Schur-convex (Schur-concave) if the indicator function $I_A(\mathbf{x})$ is Schur-convex (Schur-concave).

THEOREM 2.2. *The following statements are equivalent*:

(i) $\mathbf{X} \geqq^{\mathrm{st.m.}} \mathbf{X}'$.

(ii) $Ef(\mathbf{X}) \geqq Ef(\mathbf{X}')$ *for every Schur-convex function* f *for which both these expectations exist.*

(iii) $Ef(\mathbf{X}) \geqq Ef(\mathbf{X}')$ *for every bounded Schur-convex function* f.

(iv) $P(A) \geqq P'(A)$ *for all measurable Schur-convex sets.*

PROOF. The implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) are trivial. The implication (iv) implies (ii) follows from the fact that if f is Schur-convex and $Ef(\mathbf{X})$ and $Ef(\mathbf{X}')$ exist, then $f(\mathbf{X})$ ($f(\mathbf{X}')$) may be approximated in the $L_1(P)$ ($L_1(P')$) norm by a positive linear combination of indicator functions of Schur-convex sets. This is a consequence of the fact that for any t, $\{\mathbf{x} : f(\mathbf{x}) > t\}$ is a Schur-convex set. Finally, the implication (ii) $\Rightarrow$ (i) follows from the fact that a nondecreasing function of a Schur-convex function is Schur-convex. □

For any point $\mathbf{x} = (x_1, \cdots, x_n)$ in R_n, define $x_{[1]} \geqq \cdots \geqq x_{[n]}$ to be a nonincreasing rearrangement of $x_1, \cdots, x_n$ and define the map T from R_n into R_n by $T(\mathbf{x}) = (y_1, \cdots, y_n)$, where $y_i = \sum_{j=1}^i x_{[j]}$, $i = 1, \cdots, n$. That is, T yields the partial sums of the reverse order statistics of $x_1, \cdots, x_n$. Let $TR_n = C$. We say that a function g defined on C is *nondecreasing for each fixed* nth *coordinate* if $y_1 \geqq y_1', \cdots, y_{n-1} \geqq y_{n-1}', y_n = y_n'$ implies $g(\mathbf{y}) \geqq g(\mathbf{y}')$. Let $\mathscr{G}(n)$ denote the class of all such functions g which are Borel measurable and which are nondecreasing for each fixed nth coordinate, i.e., which are nondecreasing in each of the first $n - 1$ coordinates separately.

The following characterization of Schur-convex functions will prove useful in the characterization of stochastic majorization developed in Theorem 2.4 below.

LEMMA 2.3. *For any permutation invariant function* f *on* R_n, *define the function* g *on* C *by putting* $g(\mathbf{y}) = f(\mathbf{x})$ *whenever* $\mathbf{y} = T\mathbf{x}$. *This defines a* 1–1 *correspondence,* $f \leftrightarrow g$, *between permutation invariant functions on* R_n *and functions on* C. *Moreover,* f *is Schur-convex if and only if* g *is nondecreasing for each fixed* nth *coordinate.*

PROOF. The first statement is straightforward. The second statement is an immediate consequence of the fact that $\mathbf{x} \geqq^{\mathrm{m}} \mathbf{x}'$ if and only if $y_1 \geqq y_1', \cdots, y_{n-1} \geqq y_{n-1}', y_n = y_n'$, where $\mathbf{y} = T\mathbf{x}$ and $\mathbf{y}' = T\mathbf{x}'$. □

DEFINITION. Let $\mathbf{Y}$ and $\mathbf{Y}'$ be random vectors taking values in C. We say that $\mathbf{Y}$ is *stochastically larger* than $\mathbf{Y}'$ *for each fixed* nth *coordinate* if $g(\mathbf{Y}) \geqq^{\mathrm{st}} g(\mathbf{Y}')$ for each g in $\mathscr{G}(n)$; in symbols, $\mathbf{Y} \geqq^{\mathrm{st}} \mathbf{Y}'$ for each fixed nth coordinate.

DEFINITION. The random vector $\mathbf{X}$ is said to be stochastically larger than the random vector $\mathbf{X}'$ if $f(\mathbf{X}) \geqq^{\mathrm{st}} f(\mathbf{X}')$ for all nondecreasing functions f; we write $\mathbf{X} \geqq^{\mathrm{st}} \mathbf{X}'$.

THEOREM 2.4. *Let* $\mathbf{X}$ *and* $\mathbf{X}'$ *be random vectors in* R_n. *Set* $\mathbf{Y} = T\mathbf{X}$ *and* $\mathbf{Y}' = T\mathbf{X}'$. *Then* $\mathbf{X} \geqq^{\mathrm{st.m.}} \mathbf{X}'$ *if and only if* $\mathbf{Y} \geqq^{\mathrm{st}} \mathbf{Y}'$ *for each fixed* nth *coordinate.*

PROOF. This theorem is an immediate consequence of Lemma 2.3. □

COROLLARY 2.5. *Let* $\mathbf{X} \geqq^{\mathrm{st.m.}} \mathbf{X}'$ *and let* $\mathbf{Y} = T\mathbf{X}$, $\mathbf{Y}' = T\mathbf{X}'$. *Then* $\mathbf{Y} \geqq^{\mathrm{st}} \mathbf{Y}'$.

COROLLARY 2.6. *Let* $\mathbf{X} \geqq^{\mathrm{st.m.}} \mathbf{X}'$, $Y = T\mathbf{X}$, *and* $Y' = T\mathbf{X}'$. *Then* $Y_1 \geqq^{\mathrm{st}} Y_1', \cdots, Y_{n-1} \geqq^{\mathrm{st}} Y_{n-1}'$, *and* $Y_n =^{\mathrm{st}} Y_n'$.

COROLLARY 2.7. *Let* $\mathbf{X} \geqq^{\mathrm{st.m.}} \mathbf{X}'$ *and* $\mathbf{X}' \geqq^{\mathrm{st.m.}} \mathbf{X}$. *Let* $\mathbf{X}_{[\ ]} = (X_{[1]}, \cdots, X_{[n]})$ *and* $\mathbf{X}'_{[\ ]} = (X'_{[1]}, \cdots, X'_{[n]})$ *be the vectors of the reverse order statistics of* $\mathbf{X}$ *and* $\mathbf{X}'$, *respectively. Then* $\mathbf{X}_{[\ ]} =^{\mathrm{st}} X'_{[\ ]}$.

PROOF. Let $\mathbf{Y} = T\mathbf{X}$ and $\mathbf{Y}' = T\mathbf{X}'$. From Corollary 2.5, $\mathbf{Y} \geqq^{\mathrm{st}} \mathbf{Y}'$ and $\mathbf{Y}' \geqq^{\mathrm{st}} \mathbf{Y}$. Thus $\mathbf{Y} =^{\mathrm{st}} \mathbf{Y}'$ and $\mathbf{X}_{[\ ]} =^{\mathrm{st}} \mathbf{X}'_{[\ ]}$. □

It can be easily seen that we *cannot* have the conclusion $\mathbf{X} =^{\mathrm{st}} \mathbf{X}'$ in Corollary 2.6; for instance, let $P[\mathbf{X} = (1, 0)] = 1$, and $P[\mathbf{X}' = (1, 0)] = P[\mathbf{X}' = (0, 1)] = \frac{1}{2}$.

COUNTEREXAMPLE TO CONVERSES OF COROLLARIES 2.5 AND 2.6. Put $P[\mathbf{X} = (4, 2)] = P[\mathbf{X} = (3, 1)] = \frac{1}{2}$ and $P[\mathbf{X}' = (4, 0)] = P[\mathbf{X}' = (3, 3)] = \frac{1}{2}$. Then $Y_1 =^{\mathrm{st}} Y_1'$ and $Y_2 =^{\mathrm{st}} Y_2'$, however $\mathbf{X} \geqq^{\mathrm{st.m.}} \mathbf{X}'$ and $\mathbf{X}' \geqq^{\mathrm{st.m.}} \mathbf{X}$ are both false. This example appears in Marshall and Olkin (forthcoming).

The following statement is equivalent to the definition of majorization in the deterministic case.

EQUIVALENT DEFINITION. A vector $\mathbf{x}$ *majorizes* a vector $\mathbf{x}'$ if $\mathbf{x} \geqq^{\mathrm{m}} \mathbf{z}$ for every $\mathbf{z}$ such that $\mathbf{x}' \geqq^{\mathrm{m}} \mathbf{z}$.

A stochastic analogue of this definition (see below) will be shown to be not equivalent to our definition of stochastic majorization. The converse to Corollary 2.8 is false (via counterexample).

COROLLARY 2.8. *Let* $X \geqq^{\mathrm{st.m.}} \mathbf{X}'$. *Then for every* $\mathbf{z}$, $P[\mathbf{X} \geqq^{\mathrm{m}} \mathbf{z}] \geqq P[\mathbf{X}' \geqq^{\mathrm{m}} \mathbf{z}]$.

PROOF. The corollary follows from Theorem 2.2 (iv) and the fact that $\{\mathbf{x} : \mathbf{x} \geqq^{\mathrm{m}} \mathbf{z}\}$ is a Schur-convex set. □

Next we present a characterization of stochastic majorization which will prove to be very fruitful for producing the applications of Section 4.

THEOREM 2.9. *Let* $\mathbf{X}$ *and* $\mathbf{X}'$ *be two random vectors,* $S = \sum_1^n X_i$, *and* $S' = \sum_1^n X_i'$. *Then* $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$ *if and only if* (a) $S =^{\text{st}} S'$ *and* (b) *for each bounded Schur-convex function* f, $E(f(\mathbf{X}) \mid S = s) \geqq E(f(\mathbf{X}') \mid S' = s)$ *for all* $s \in A_f$, *where* A_f *satisfies* $P[S \in A_f] = 1$.

PROOF. First let $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$. Then (a) follows from Lemma 2.1.

To prove (b), consider the function $g(\mathbf{x}) = f(\mathbf{x})I(s \in A)$, where f is bounded Schur-convex, $s = \sum_1^n x_i$, A is a Borel set in R_1, and $I(s \in A) = 1$ if $s \in A$ and $= 0$ otherwise. Then g is a Schur-convex function. Thus $E[f(\mathbf{X})I(S \in A)] \geqq E[f(\mathbf{X}')I(S' \in A)]$ for each Borel set A in R_1. Since $S =^{\text{st}} S'$, it follows that $E[f(\mathbf{X}) \mid S = s] \geqq E[f(\mathbf{X}') \mid S' = s]$ for each $s \in A_f$ satisfying $P[S \in A_f] = 1$.

The converse follows simply by unconditioning using the common distribution of S and S'. □

As the notation indicates, the set A_f, in general, depends on the particular Schur function f. However, in many applications this is no disadvantage since one often wishes to establish an inequality for one single Schur function at a time and the conclusion of Theorem 2.9 is strong enough for this purpose.

If the A_f appearing in Theorem 2.9 could be chosen to be independent of f, then we could claim that the conditional distribution of $\mathbf{X}$ given $S = s$ stochastically majorizes the conditional distribution of $\mathbf{X}'$ given $S' = s$ for almost all s with respect to the common distribution of S and S'. This, of course, would be a stronger and neater conclusion than that of Theorem 2.9. This stronger version will be established below in Corollary 2.10 when $\mathbf{X}$ and $\mathbf{X}'$ have discrete distributions. The dependence of A_f on f in Theorem 2.9 can be eliminated if the class of Schur-convex functions needed to check stochastic majorization can be reduced to a countable class of functions. The following conjecture expresses this idea in terms of Schur-convex sets.

CONJECTURE. There exists a countable collection $\mathscr{D}$ of Schur-convex sets such that for any Schur-convex set F, any $\varepsilon > 0$, and any probability measure μ, there is a member $D = D(F, \mu, \varepsilon)$ of $\mathscr{D}$ such that $\mu[(F \cap D^c) \cup (D \cap F^c)] < \varepsilon$.

By relating Schur-convex sets to increasing sets and using the discussion following Lemma 2 in Blum (1955), we have been able to show that a countable class $\mathscr{D}$ of Schur-convex sets as mentioned in the above conjecture exists that can be used for all absolutely continuous and all discrete probability measures μ. Since this represents only a partial answer, we do not present these results here.

COROLLARY 2.10. *Let* $\mathbf{X}$ *and* $\mathbf{X}'$ *be random vectors. Let* $S = \sum_1^n X_i$ *and* $S' = \sum_1^n X_i'$, *where the distribution of* S *is discrete. Finally, let* $K = \{s : P[S = s] > 0\}$. *Then* $X \geqq^{\text{st.m.}} \mathbf{X}'$ *if and only if* (a) $S =^{\text{st}} S'$ *and* (b) *the conditional distribution of* $\mathbf{X}$ *given* $S = s$ *stochastically majorizes the conditional distribution of* $\mathbf{X}'$ *given* $S' = s$ *for each* s *in* K.

PROOF. The proof is immediate from Theorem 2.9 since the A_f appearing in Theorem 2.9 must include K. □

Note that Corollary 2.10 yields a stronger conclusion than does Theorem 2.9 under the assumption that S has a discrete distribution.

3. Preservation of stochastic majorization under operations. The extent of applicability of our new notion of stochastic majorization is dependent on the degree to which it is preserved under various standard mathematical, probabilistic, and statistical operations. In this section we display operations which preserve stochastic majorization.

First we show that stochastic majorization is preserved under mixtures of distributions.

THEOREM 3.1. *Let $\mathbf{X}$ and $\mathbf{X}'$ be two random vectors and let U be a random variable such that the conditional distribution of $\mathbf{X}$ given $U = u$ stochastically majorizes the conditional distribution of $\mathbf{X}'$ given $U = u$ for each u. Then $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$.*

PROOF. Let $g(\mathbf{x})$ be a bounded Schur-convex function. Then by Theorem 2.2, $Eg(\mathbf{X}) = EEg(\mathbf{X} \mid U) \geqq EEg(\mathbf{X}' \mid U) = Eg(\mathbf{X}')$. Again by Theorem 2.2, $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$. □

Next, we show that stochastic majorization is preserved under a normalization operation. The result will be used to generate a number of applications in Section 4.

THEOREM 3.2. *Let $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$ and f be a Borel-measurable function on R_1. Then*

$$f(\textstyle\sum_1^n X_i)\mathbf{X} \geqq^{\text{st.m.}} f(\sum_1^n X_i')\mathbf{X}' .$$

PROOF. The result follows from the fact that if $g(\mathbf{x})$ is a Schur-convex function of $\mathbf{x}$, then so is $g(f(\sum_1^n x_i)\mathbf{x})$. □

The most important and useful operation which preserves stochastic majorization is presented in the following theorem.

THEOREM 3.3 (Preservation Theorem). *Let X_{λ_i} have density $\phi(\lambda_i, x)$ with respect to Lebesgue measure or to counting measure. Let $\phi(\lambda, x) = 0$ for $x < 0$ and let $\phi(\lambda, x)$ be TP_2 for $0 < \lambda < \infty$, $x \geqq 0$ and satisfy the semigroup property for $0 < \lambda < \infty$. Let $\mathbf{X}_\lambda = (X_{\lambda_1}, \cdots, X_{\lambda_n})$ be a random vector of independent components. Then $\boldsymbol{\lambda} \geqq^{\text{m}} \boldsymbol{\lambda}'$ implies that $X_\lambda \geqq^{\text{st.m.}} \mathbf{X}_{\lambda'}$.*

PROOF. The result follows directly from Theorem 1.1 of Part I and Theorem 2.2 above. □

Thus a deterministic property (majorization) of the parameter vector $\boldsymbol{\lambda}$ is transformed into a corresponding stochastic property (stochastic majorization) of the random vector $\mathbf{X}_\lambda$. This leads to the definition of Schur families of random vectors and of multivariate distributions.

DEFINITION. Let X_λ be a random vector with distribution P_λ in R^n indexed by a parameter vector $\boldsymbol{\lambda}$ in R^k. The family $\{\mathbf{X}_\lambda\}$ of random vectors and the corresponding family $\{P_\lambda\}$ of multivariate distributions are said to form *Schur families* in $\boldsymbol{\lambda}$ if $\boldsymbol{\lambda} \geqq^{\text{m}} \boldsymbol{\lambda}'$ implies $\mathbf{X}_\lambda \geqq^{\text{st.m.}} \mathbf{X}_{\lambda'}$.

Theorem 3.3 shows how to obtain Schur families of multivariate distributions. We exploit this theorem in Section 4.

An immediate corollary of Theorem 3.3 is an application to stochastic processes.

COROLLARY 3.4. *Let $\{X(t), 0 \leqq t < \infty\}$ be a stochastic process with stationary, independent, and nonnegative increments. Let the density $\phi(\lambda, x)$ of $X(t+\lambda) - X(t)$, with respect to Lebesgue measure or counting measure, be TP_2 in $\lambda > 0$ and $x \geqq 0$. Let $0 = t_0 \leqq t_1 \leqq \cdots \leqq t_n = T$, $0 = t_0' \leqq t_1' \leqq \cdots \leqq t_n' = T$, $\lambda_i = t_i - t_{i-1}$, and $\lambda_i' = t_i' - t_{i-1}'$, $i = 1, \cdots, n$. Then $\boldsymbol{\lambda} \geqq^{\text{m}} \boldsymbol{\lambda}'$ implies*

$$(X(t_1) - X(t_0), \cdots, X(t_n) - X(t_{n-1})) \geqq^{\text{st.m.}} (X(t_1') - X(t_0'), \cdots, X(t_n') - X(t_{n-1}')) .$$

PROOF. Note that the semigroup property holds because the increments are stationary and independent. The result follows by an application of Theorem 3.3. □.

The next theorem shows that stochastic majorization is preserved after an introduction of prior distributions if there is stochastic majorization among the prior distributions.

THEOREM 3.5. *Let $\{\mathbf{X}_\lambda\}$ be a Schur family in $\boldsymbol{\lambda}$. Let G_1, G_2 be two probability measures on R_k such that $G_1 \geqq^{\text{st.m.}} G_2$. Let $Q_i(A) = \int_{R_k} P(\mathbf{X}_\lambda \in A)\, dG_i(\boldsymbol{\lambda})$ for all Borel sets A in R_k, $i = 1, 2$. Then $Q_1 \geqq^{\text{st.m.}} Q_2$.*

PROOF. Let $g(\mathbf{x})$ be a bounded Schur-convex function. Then,

$$\int g(\mathbf{x})\, dQ_1(\mathbf{x}) = \int E[g(\mathbf{X}_\lambda)]\, dG_1(\boldsymbol{\lambda}) \geqq \int E[g(\mathbf{X}_\lambda)]\, dG_2(\boldsymbol{\lambda}) = \int g(\mathbf{x})\, dQ_2(\mathbf{x})$$

since $E[g(\mathbf{X}_\lambda)]$ is a Schur-convex function of $\boldsymbol{\lambda}$. Thus $Q_1 \geqq^{\text{st.m.}} Q_2$. □

One further preservation theorem will be useful in applications. The following easy theorem considers preservation of stochastic majorization under a limiting operation and is presented here without proof.

THEOREM 3.6. *Let $\{\mathbf{X}_n\}$ and $\{\mathbf{X}_n'\}$ be sequences of random vectors such that $\mathbf{X}_n \geqq^{\text{st.m.}} \mathbf{X}_n'$ for $n = 1, 2, \cdots$. Suppose $\mathbf{X}_n \to \mathbf{X}$ and $\mathbf{X}_n' \to \mathbf{X}'$ in the sense that $Ef(\mathbf{X}_n) \to Ef(\mathbf{X})$ for all bounded measurable functions f. Then $\mathbf{X} \geqq^{\text{st.m.}} \mathbf{X}'$.*

4. Applications of stochastic majorization. In this section we present applications of the new notion of stochastic majorization to obtain general inequalities for a wide variety of standard multivariate distributions.

APPLICATION 4.1. *Let $\mathbf{X}_\lambda = (X_{\lambda_1}, \cdots, X_{\lambda_n})$ be a random vector of independent components, where X_{λ_i}, $i = 1, \cdots, n$, has a density of the form given in* (a), (b), *or* (c) *below. Then the family of random vectors $\{\mathbf{X}_\lambda\}$ is a Schur family in $\boldsymbol{\lambda}$.*

(a) Poisson. $\phi(\lambda, x) = (\lambda\theta)^x e^{-\lambda\theta}/x!$, $x = 0, 1, \cdots, \lambda > 0$, *and fixed* $\theta > 0$.

(b) Binomial. $\phi(\lambda, x) = \binom{\lambda}{x} p^x(1-p)^{\lambda-x}$, $x = 0, 1, \cdots, \lambda$; $\lambda = 1, 2, \cdots$; *and fixed* p *with* $0 < p < 1$.

(c) Gamma. $\phi(\lambda, x) = (\theta^\lambda x^{\lambda-1}/\Gamma(\lambda))e^{-\theta\lambda}$, $x \geqq 0$, $\lambda > 0$, *and fixed* $\theta > 0$.

PROOF. The result follows from Theorem 3.3 by noting that ϕ is TP_2 and satisfies the semigroup property in each case (a), (b), and (c). □

Application 4.1 may be used to obtain stochastic majorization comparisons for the multinomial, multivariate negative binomial, multivariate hypergeometric, multivariate negative hypergeometric, Dirichlet and multivariate logarithmic series distributions by using the preservation results of Section 3. The following summarizes these stochastic majorization comparisons.

APPLICATION 4.2. *Let* $\mathbf{Y}_\lambda = (Y_{1,\lambda}, \cdots, Y_{n,\lambda})$ *have any one of the distributions specified in* (a), (b), (c), (d), (e), *or* (f) *below. Then the family of random vectors* $\{\mathbf{Y}_\lambda\}$ *is a Schur family in* $\boldsymbol{\lambda}$.

(a) Multinomial. $f_\lambda(\mathbf{y}) = N! \prod_{i=1}^n (\lambda_i^{y_i}/y_i!)$, *where* $y_i = 0, 1, \cdots, n$, $\sum_{i=1}^n y_i = N$; $\lambda_i > 0$, $i = 1, \cdots, n$, $\sum_{i=1}^n \lambda_i = 1$.

(b) Multivariate negative binomial.

$$f_\lambda(\mathbf{y}) = \frac{\Gamma(N + \sum_{i=1}^n y_i)}{\Gamma(N)} \prod_{i=1}^n \frac{\lambda_i^{y_i}}{y_i!} [1 + \sum_{i=1}^n \lambda_i]^{-N-\sum_{i=1}^n y_i}$$

where $y_i = 0, 1, \cdots, i = 1, \cdots, n$; $\lambda_i > 0$, $i = 1, \cdots, n$, $N > 0$.

(c) Multivariate hypergeometric.

$$f_\lambda(\mathbf{y}) = \frac{\prod_{i=1}^n \binom{\lambda_i}{y_i}}{\binom{\sum_{i=1}^n \lambda_i}{N}},$$

where $y_1, \cdots, y_n$ *are integers satisfying* $0 \leqq y_i \leqq \lambda_i$, $i = 1, \cdots, n$, *and* $\sum_1^n y_i = N$, $0 < N \leqq \sum_1^n \lambda_i$, *and* $\lambda_1, \cdots, \lambda_n$ *are positive integers*.

(d) Dirichlet. $\mathbf{Y}_\lambda =^{\text{st}} (1/\sum_1^n X_{\lambda_i})\mathbf{X}_\lambda$, *where* $\mathbf{X}_\lambda = (X_{\lambda_1}, \cdots, X_{\lambda_n})$ *is a vector of independent gamma random variables as in Application* 4.1 (c).

(e) Negative multivariate hypergeometric.

$$f_\lambda(\mathbf{y}) = \frac{N!\,\Gamma(\sum_{j=1}^n \lambda_j)}{\prod_{j=1}^n y_j!\,\Gamma(N + \sum_{j=1}^n \lambda_j)} \prod_{j=1}^n \frac{\Gamma(y_j + \lambda_j)}{\Gamma(\lambda_j)},$$

where $y_j = 0, 1, \cdots, N$, $\sum_{i=1}^n y_i = N$, $\lambda_i > 0$, $i = 1, \cdots, n$.

(f) Multivariate logarithmic series.

$$f_\lambda(\mathbf{y}) = \frac{(\sum_{i=1}^n y_i - 1)!}{\log(1 + \sum_{i=1}^n \lambda_i)} \prod_{i=1}^n \frac{\lambda_i^{y_i}}{y_i!} (1 + \sum_{i=1}^n \lambda_i)^{-\sum_{i=1}^n y_i}$$

where $y_i = 0, 1, \cdots, \lambda_i > 0$, $i = 1, \cdots, n$, *and* $\sum_{i=1}^n y_i > 0$..

PROOF.

(a) This application follows from Applications 4.1 (a), Corollary 2.10, and the fact that the multinomial distribution is the conditional distribution of independent Poisson random variables given their sum.

(b) This application follows from Application 4.1 (a), Theorem 3.1, and the fact that the multivariate negative binomial distribution is a mixture of independent Poisson random variables under a gamma distribution (Johnson and Kotz (1969), page 293).

(c) This application follows from Corollary 2.10 and the fact that the multivariate hypergeometric distribution is the conditional distribution of independent binomials given their sum.

(d) This application follows from Application 4.1 (c) and Theorem 3.2.

(e) This application follows from (a), Theorem 3.5, and the fact that a mixture of a multinomial distribution with a Dirichlet distribution for the parameter vector $\boldsymbol{\lambda}$ is a multivariate negative hypergeometric distribution (Johnson and Kotz (1969), page 309).

(f) The multivariate logarithmic series distribution is the limit of the conditional distribution of a multivariate negative binomial given that the sum is positive as the parameter N goes to zero (Johnson and Kotz (1969), page 302). Application (f) follows from this fact, Application 4.2 (b) above, and Theorem 3.6. □

We may use Application 4.2 to prove a generalized version of Lemma 1.2 of Part I.

APPLICATION 4.3. *Let* $\mathbf{Y}_\lambda$ *have any one of the densities specified in* 4.2 (a), (b), (c), (e), *or* (f). *Let* Z_λ *be the number of zero components of the vector* $\mathbf{Y}_\lambda$. *Then* $\boldsymbol{\lambda} \geqq^{\mathrm{m}} \boldsymbol{\lambda}'$ *implies* $Z_\lambda \geqq^{\mathrm{st}} Z_{\lambda'}$.

PROOF. The desired conclusion follows from Application 4.2, Theorem 2.2 (iii) and the fact that $g(\mathbf{y}) = \sum_1^n I_{[y_i=0]}$ is Schur-convex for $y_1 \geqq 0, \cdots, y_n \geqq 0$. □

Lemma 1.2 of Part I follows directly from Application 4.3.

In the next application, we generalize a result of Olkin (1972).

APPLICATION 4.4. *Let* $\mathbf{Y}_\lambda$ *have any one of the densities specified in Application* 4.2 (a), (b), (c), (d), (e), *or* (f). *Then* $\boldsymbol{\lambda} \geqq^{\mathrm{m}} \boldsymbol{\lambda}'$ *implies*

$$P[Y_{1,\lambda} \leqq y, \cdots, Y_{n,\lambda} \leqq y] \leqq P[Y_{1,\lambda'} \leqq y, \cdots, Y_{n,\lambda'} \leqq y] \tag{4.1}$$

and

$$P[Y_{1,\lambda} > y, \cdots, Y_{n,\lambda} > y] \leqq P[Y_{1,\lambda'} > y, \cdots, Y_{n,\lambda'} > y] \tag{4.2}$$

for all y.

PROOF. The desired conclusions (4.1) and (4.2) follow from Theorem 2.2 (iv) and the fact that $\{\mathbf{x} : x_i > y,\ i = 1, \cdots, n\}$ and $\{\mathbf{x} : x_i \leqq y,\ i = 1, \cdots, n\}$ are Schur-concave. □

A result in Olkin (1972) corresponds to the multinomial case (4.2 (a)).

We will now show that the absolutely continuous Dirichlet distributions form a Schur family. Let $\theta > 0$ be fixed throughout our next application. Let Z_λ be a gamma random variable with density $\phi(\lambda, x) = (x^{\lambda-1}/\Gamma(\lambda))e^{-x}$ for $x \geqq 0$, $\lambda > 0$. Let $Z_{\lambda_1}, \cdots, Z_{\lambda_n}, Z_\theta$ be mutually independent. Let $\mathbf{Y}_\lambda$ be an n-variate random vector with components $Y_{i,\lambda} = Z_{\lambda_i}/(Z_\theta + \sum_1^n Z_{\lambda_i})$, $i = 1, \cdots, n$. As pointed out in Wilks (1962), page 179, $\mathbf{Y}_\lambda$ has an n-variate absolutely continuous Dirichlet distribution ($D(\lambda_1, \cdots, \lambda_n; \theta)$, in symbols), with density

$$f(\mathbf{y}) = \frac{\Gamma(\theta + \sum_1^n \lambda_i)}{\Gamma(\theta) \prod_1^n \Gamma(\lambda_i)} (1 - \sum_1^n y_i)^{\theta-1} \prod_1^n y_i^{\lambda_i - 1} \tag{4.3}$$

for $y_i \geqq 0$, $i = 1, \cdots, n$, and $\sum_1^n y_i \leqq 1$.

APPLICATION 4.5. *The family of Dirichlet distributions with density given in* (4.3) *forms a Schur family in* $\boldsymbol{\lambda}$.

PROOF. Conditional on $Z_\theta = z$, we may write $\mathbf{Y}_\lambda =^{\mathrm{st}} (Z_{\lambda_1}/(z + S), \cdots, Z_{\lambda_n}/(z + S))$ where $S = \sum_1^n Z_{\lambda_i}$. From Application 4.1 (c) and Theorem 3.2, the conditional distribution of $\mathbf{Y}_\lambda$ given $Z_\theta = z$ stochastically majorizes the conditional distribution of $\mathbf{Y}_{\lambda'}$ given $Z_\theta = z$. The present application now follows from Theorem 3.1. □

A useful application of stochastic majorization for Dirichlet distributions is next developed for the coverages from a continuous distribution. Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics in a sample of size n from a continuous distribution F. Define coverages $U_1 = F(X_{(1)})$, $U_2 = F(X_{(2)}) - F(X_{(1)})$, $\cdots$, $U_n = F(X_{(n)}) - F(X_{(n-1)})$. Let V_i denote the sum of r_i of these coverages, $i = 1, \cdots, k$, with no coverage belonging to more than one V_i. Then as pointed out by Wilks (1962), page 238, $\mathbf{V}_\mathbf{r} = (V_1, \cdots, V_k)$ has an absolutely continuous Dirichlet distribution, $D(r_1, \cdots, r_k; n + 1 - \sum_1^k r_i)$. Using Application 4.5 we obtain:

APPLICATION 4.6. *Let* $V_1, \cdots, V_k$ *be sums of distinct coverages as specified just above. Then* $\mathbf{r} \geqq^{\mathrm{m}} \mathbf{r}'$ *implies* $\mathbf{V}_\mathbf{r} \geqq^{\mathrm{st.m.}} \mathbf{V}_{\mathbf{r}'}$.

REMARK. If a $(k + 1)$st component $V_{k+1} = 1 - \sum_1^k V_i$ is added to the vector $\mathbf{V}_\mathbf{r}$ above, then the resulting vector $\mathbf{V}_\mathbf{r}^{(s)} = (V_1, \cdots, V_k, V_{k+1})$ has the singular Dirichlet distribution of Application 4.2 (d). It follows that $(r_1, \cdots, r_k, n + 1 - \sum_1^k r_i) \geqq^{\mathrm{m}} (r_1', \cdots, r_k', n + 1 - \sum_1^k r_i')$ implies that $\mathbf{V}_\mathbf{r}^{(s)} \geqq^{\mathrm{st.m.}} \mathbf{V}_{\mathbf{r}'}^{(s)}$.

APPLICATION 4.7. *Let* $\mathbf{Y}_\lambda$ *have an inverted Dirichlet distribution with density*

$$f_\lambda(\mathbf{y}) = \frac{\Gamma(\theta + \sum_1^n \lambda_i)}{\Gamma(\theta) \prod_1^n \Gamma(\lambda_i)} \frac{\prod_1^n y_j^{\lambda_j - 1}}{(1 + \sum_1^n y_j)^{\theta + \sum_1^n \lambda_j}}$$

$$\text{for } y_i \geqq 0\,, \quad \lambda_i > 0\,, \quad i = 1, \cdots n\,,$$

and fixed $\theta > 0$. (See Johnson and Kotz, 1972, page 239.) *Then the family of random vectors* $\{Y_\lambda\}$ *is a Schur family in* $\boldsymbol{\lambda}$.

PROOF. Let $Z_{\lambda_1}, \cdots, Z_{\lambda_n}, Z_\theta$ denote independent gamma random variables with parameters $\lambda_1, \cdots, \lambda_n, \theta$, respectively. Then one can write $\mathbf{Y}_\lambda =^{st} (Z_{\lambda_1}/Z_\theta, \cdots, Z_{\lambda_n}/Z_\theta)$. The present application follows from Application 4.1 (c) and Theorem 3.1 by first conditioning on Z_θ and then unconditioning. □

The applications listed so far yield examples of stochastic majorization among familiar distributions by repeated applications of Theorems 2.2, 3.1, 3.2, 3.3, 3.5 and Corollary 2.10.

The next application is a reformulation of Corollary 3.3 of Part I and is stated here so that we may, by means of an example, show that in the preservation theorem (Theorem 1.1 of Part I), from which most of the applications follow, the TP_2 assumption cannot be dispensed with, in general.

APPLICATION 4.8. *Let* X_{ij}, $i = 1, \cdots, n$; $j = 1, 2, \cdots$, *be independently and identically distributed according to a* log *concave density* g *with support* $[0, \infty)$. *Let* $S_{i,k_i} = \sum_{j=1}^{k_i} X_{ij}$, $i = 1, \cdots, n$, *and* $\mathbf{S}_\mathbf{k} = (S_{1,k_1}, \cdots, S_{n,k_n})$. *Then* $\mathbf{k} \geqq^{m} \mathbf{k}'$ *implies that* $\mathbf{S}_\mathbf{k} \geqq^{\text{st.m.}} \mathbf{S}_{\mathbf{k}'}$.

Putting $n = 2$, $k_1 = 3$, $k_2 = 1$, $k_1' = 2$, $k_2' = 2$ and choosing the Schur-convex function $g(x_1, x_2) = x_1^2 + x_2^2$, we have $(X_1 + X_2 + X_3)^2 + X_4^2 \geqq^{st} (X_1 + X_2^2) + (X_3 + X_4)^2$ as a special case of Application 4.8, whenever X_1, X_2, X_3 and X_4 are i.i.d. nonnegative random variables with a log-concave density. Even this special case is false, in general, without the assumption of a log-concave density as can be seen from the example given below. Let $P[X_i = 1] = \frac{1}{2}$, $P[X_i = 3] = \frac{1}{2}$. It can be checked that the frequency function of X_i is *not* log-concave and $(X_1 + X_2 + X_3)^2 + X_4^2$ is *not* stochastically larger than $(X_1 + X_2)^2 + (X_3 + X_4)^2$, since $P[(X_1 + X_2 + X_3)^2 + X_4^2 > 19] = \frac{7}{8}$, while $P[(X_1 + X_2)^2 + (X_3 + X_4)^2 > 19] = \frac{15}{16}$.

In fact the above example can be used to show that the condition $\phi(\lambda, x)$ be TP_2 cannot, in general, be dropped from Theorem 1.1 of Part I. To see this, put $\phi(n, x) = P(X_1 + \cdots + X_n = x)$, $n = 1, 2, \cdots$ where the X_i's are as in the example given above. Then $\phi(n, x)$ satisfies the semigroup property but is not TP_2 since $\begin{vmatrix}\phi(1,2) & \phi(2,2)\\ \phi(1,3) & \phi(2,3)\end{vmatrix} = \begin{vmatrix}0 & \frac{1}{4}\\ \frac{1}{2} & 0\end{vmatrix} < 0$. The conclusion of Theorem 1.1 of Part I does not hold for the Schur-convex function $g(x_1, x_2)$ defined by $g(x_1, x_2) = 1$, if $x_1^2 + x_2^2 > 19$ and $= 0$ otherwise, since,

$$\begin{aligned}\tfrac{7}{8} &= \textstyle\sum_{x_1,x_2} \phi(3, x_1)\phi(1, x_1)g(x_1, x_2) = P[(X_1 + X_2 + X_3)^2 + X_4^2 > 19] \\ &< \textstyle\sum_{x_1,x_2} \phi(2, x_1)\phi(2, x_1)g(x_1, x_2) = P[(X_1 + X_2)^2 + (X_3 + X_4)^2 > 19] = \tfrac{15}{16}\,.\end{aligned}$$

REFERENCES

[1] BLUM, J. R. (1955). On the convergence of empiric distribution functions. *Ann. Math. Statist.* **26** 527–529.

[2] JOHNSON, N. L. and KOTZ, S. (1969). *Distributions in Statistics: Discrete Distributions.* Wiley, New York.

[3] JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

[4] MARSHALL, A. W. and OLKIN, I. (Forthcoming). *Majorization and Schur Functions.* Academic Press, New York.

[5] OLKIN, I. (1972). Monotonicity properties of Dirichlet integrals with applications to the multinomial distribution and the analysis of variance. *Biometrika* **59** 303–307.
[6] WILKS, S. S. (1962). *Mathematical Statistics.* Wiley, New York.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306