The Annals of Statistics
1982, Vol. 10, No. 2, 637–642

RECTANGULAR AND ELLIPTICAL PROBABILITY INEQUALITIES FOR SCHUR-CONCAVE RANDOM VARIABLES[1]

BY Y. L. TONG

University of Nebraska

It is shown that if the density $f(\mathbf{x})$ of $\mathbf{X} = (X_1, \cdots, X_n)$ is Schur-concave, then (1) $P(|X_i| \le a_i,\ i = 1, \cdots, n)$ is a Schur-concave function of $(\phi(a_1), \cdots, \phi(a_n))$, and (2) $P\{\Sigma(X_i/a_i)^2 \le 1\}$ is a Schur-concave function of $(\phi(a_1^2), \cdots, \phi(a_n^2))$, where $\phi; [0, \infty) \to [0, \infty)$ is any increasing and convex function. By letting $\phi(a) = a$, (1) implies that $P(|X_i| \le a_i,\ i = 1, \cdots, n) \le P(|X_i| \le \bar{a},\ i = 1, \cdots, n)$. As special consequences, the results yield bounds for exchangeable normal and t variables and for linear combinations of central and noncentral Chi squared variables.

1. Introduction and motivation. This work concerns probability inequalities via majorization and Schur-concavity. Roughly speaking, $\mathbf{a}$ majorizes $\mathbf{b}$ (in symbols, $\mathbf{a} \succ \mathbf{b}$) indicates that the components of $\mathbf{a}$ are more diverse. For definitions of majorization and Schur-concave functions, see Marshall and Olkin (1979, page 7, page 54).

Let $\mathbf{X} = (X_1, \cdots, X_n)$ denote a n-dimensional random variable whose density is $f(\mathbf{x})$. We shall say that $\mathbf{X}$ is Schur-concave if $f(\mathbf{x})$ is a Schur-concave function of $\mathbf{x}$. Marshall and Olkin (1974) previously proved that the cumulative probability

$$\beta(\mathbf{a}) \equiv P(X_i \le a_i,\ i = 1, \cdots, n) \tag{1.1}$$

is a Schur-concave function of $\mathbf{a} = (a_1, \cdots, a_n)$ for all Schur-concave random variables $\mathbf{X}$. Since in many statistical applications the rectangular probability

$$\gamma(\mathbf{a}) \equiv P(|X_i| \le a_i,\ i = 1, \cdots, n) \tag{1.2}$$

is also of concern, a related question is whether or not $\gamma(\mathbf{a})$ is also a Schur-concave function of $\mathbf{a}$. This question cannot be answered by a similar argument given by Marshall and Olkin (1974), so a new proof is needed.

In this paper we show that the answer to the above question is positive by proving an integral inequality for Schur-concave functions. After this result is proved, we then use the same basic argument to obtain an inequality for the probability contents of ellipsoids, and indicate a direction for possible generalization. The inequalities involve the concept of diversity of the components of a scale (instead of location) parameter vector, and are applicable in a number of situations. For the multivariate normal and t variables, the inequalities yield bounds for the probability contents of rectangular sets and ellipsoids. Consequently bounds for linear combinations of central and noncentral Chi squared variables can be derived.

After proving these inequalities in Section 2, we observe some applications in Section 3.

2. The main theorems. For notational convenience let $A(\mathbf{a})$, $B(\mathbf{a})$ denote the

Received January 1981; revised October 1981.

[1] This research was partially supported by NSF Grant MCS-8100775.

AMS 1980 *subject classifications*. Primary 62H99, 26D15, 60E15.

Key words and phrases. Probability inequalities in multivariate distributions, Schur-functions and majorization, bounds for multivariate normal and t probabilities, bound for central and noncentral Chi squared probabilities.

rectangular set and the ellipsoid given by

$$A(\mathbf{a}) = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, |x_i| \le a_i, i = 1, \cdots, n\}, \tag{2.1}$$

$$B(\mathbf{a}) = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \textstyle\sum_{i=1}^n (x_i/a_i)^2 \le 1\}. \tag{2.2}$$

To avoid the trivial case, it will be assumed that the a_i's are strictly > 0 for $i = 1, \cdots, n$. The purpose of this section is to present the following theorems.

THEOREM 2.1. *If $f(\mathbf{x}) : \mathbb{R}^n \to [0, \infty)$ is Borel-measurable and Schur-concave (strictly Schur-concave) then, provided that the integral exists, $\int_{A(\mathbf{a})} f(\mathbf{x})\, d\mathbf{x}$ is a Schur-concave (strictly Schur-concave) function of $(a_1, \cdots, a_n)$.*

THEOREM 2.2. *If $f(\mathbf{x})$ satisfies the conditions stated in Theorem 2.1, then $\int_{B(\mathbf{a})} f(\mathbf{x})\, d\mathbf{x}$ is a Schur-concave (strictly Schur-concave) function of $(a_1^2, \cdots, a_n^2)$.*

The proofs of these two theorems depend on the lemma given below. Let $C(\mathbf{a})$, $C(\mathbf{b})$ denote two closed convex sets in $\mathbb{R}^2$ which depend on the vectors (a_1, a_2) and (b_1, b_2). Let

$$c_1(\mathbf{a}) = \inf\{x_1 + x_2 \mid \mathbf{x} \in C(\mathbf{a})\}, \quad c_2(\mathbf{a}) = sup\{x_1 + x_2 \mid \mathbf{x} \in C(\mathbf{a})\}.$$

For every fixed $\lambda \in [c_1(\mathbf{a}), c_2(\mathbf{a})]$ let $\ell_\lambda(\mathbf{a})$ be the segment of the line $x_1 + x_2 = \lambda$ intersecting $C(\mathbf{a})$, $\| \ell_\lambda(\mathbf{a}) \|$ the length of $\ell_\lambda(\mathbf{a})$, and $\delta_\lambda(\mathbf{a})$ the distance between the midpoint of $\ell_\lambda(\mathbf{a})$ and $(\frac{1}{2}\lambda, \frac{1}{2}\lambda)$. Let us assume the following conditions

CONDITION A1. $c_1(\mathbf{a}) \ge c_1(\mathbf{b})$ and $c_2(\mathbf{a}) \le c_2(\mathbf{b})$.

CONDITION A2. The inequalities

$$\| \ell_\lambda(\mathbf{a}) \| \le \| \ell_\lambda(\mathbf{b}) \|, \tag{2.3}$$

$$\delta_\lambda(\mathbf{a}) \ge \delta_\lambda(\mathbf{b}) \tag{2.4}$$

hold for all $\lambda \in [c_1(\mathbf{a}), c_2(\mathbf{a})]$.

Then the set $\{x_1 + x_2 \mid \mathbf{x} \in C(\mathbf{a})\}$ is a subset of $\{x_1 + x_2 \mid \mathbf{x} \in C(\mathbf{b})\}$, the line segment in $C(\mathbf{a})$ is shorter, and the coordinates of its midpoint are more diverse in the sense of majorization.

LEMMA 2.1. *Assume that Conditions* A1, A2 *are satisfied. If $f(\mathbf{x}) : \mathbb{R}^2 \to [0, \infty)$ is Borel-measurable and Schur-concave then, provided that the integrals exist,*

$$\int_{C(\mathbf{a})} f(\mathbf{x})\, d\mathbf{x} \le \int_{C(\mathbf{b})} f(\mathbf{x})\, d\mathbf{x} \tag{2.5}$$

holds. In addition, if $f(\mathbf{x})$ is strictly Schur-concave and if strict inequalities in (2.3), (2.4) *hold for $\lambda \in I_1$ and $\lambda \in I_2$, respectively for some intervals I_1 and I_2, then the inequality in* (2.5) *is strict.*

PROOF. Consider the orthogonal transformation

$$u_1 = (x_1 + x_2)/\sqrt{2}, \qquad u_2 = (x_1 - x_2)/\sqrt{2}.$$

For each fixed u_1 let $(u_1, d_1(\mathbf{a}))$ and $(u_1, d_2(\mathbf{a}))$ be the endpoints of the line segment in $C(\mathbf{a})$ in the (u_1, u_2) space, and $(u_1, d(\mathbf{a}))$ be the midpoint, where $d_1(\mathbf{a}) \le d_2(\mathbf{a})$, $d(\mathbf{a}) = \frac{1}{2} \sum_1^2 d_i(\mathbf{a})$, and they depend on u_1. Then Condition A2 implies

$$d_2(\mathbf{a}) - d_1(\mathbf{a}) \le d_2(\mathbf{b}) - d_1(\mathbf{b}), \tag{2.6}$$

$$|d(\mathbf{a})| - |d(\mathbf{b})| \ge 0. \tag{2.7}$$

If we write

$$\int_{C(\mathbf{a})} f(\mathbf{x})\, d\mathbf{x} = \int_{c_1(\mathbf{a})/\sqrt{2}}^{c_2(\mathbf{a})/\sqrt{2}} \int_{d_1(\mathbf{a})}^{d_2(\mathbf{a})} f((u_1 + u_2)/\sqrt{2}, (u_1 - u_2)/\sqrt{2})\, du_2\, du_1,$$

then for every fixed u_1, by Schur-concavity, the function $f((u_1 + u_2)/\sqrt{2}, (u_1 - u_2)/\sqrt{2})$ is symmetric in u_2 and unimodal. Thus for fixed $d_2(\mathbf{a}) - d_1(\mathbf{a})$ the value of the inner integral is nonincreasing in $|d(\mathbf{a})|$. Letting

$$\varepsilon = \begin{cases} d(\mathbf{a}) - d(\mathbf{b}) & \text{if } d(\mathbf{a}) \geq 0, \\ d(\mathbf{b}) - d(\mathbf{a}) & \text{otherwise,} \end{cases}$$

and combining (2.6), (2.7) with the fact that $f \geq 0$, we have

$$\begin{aligned} \int_{d_1(\mathbf{a})}^{d_2(\mathbf{a})} f((u_1 + u_2)/\sqrt{2}, (u_1 - u_2)/\sqrt{2})\, du_2 &\leq \int_{d_1(\mathbf{a})-\varepsilon}^{d_2(\mathbf{a})-\varepsilon} f((u_1 + u_2)/\sqrt{2}, (u_1 - u_2)/\sqrt{2})\, du_2 \\ &\leq \int_{d_1(\mathbf{b})}^{d_2(\mathbf{b})} f((u_1 + u_2)/\sqrt{2}, (u_1 - u_2)/\sqrt{2})\, du_2. \end{aligned} \tag{2.8}$$

Combining (2.8) with Condition A1, the inequality (2.5) now follows. Moreover, the inequalities in (2.8) are strict, hence (2.5) is strict, under the additional conditions. □

PROOF OF THEOREM 2.1. In view of Lemma B.1 on page 21 of Marshall and Olkin (1979) we can give the proof for $n = 2$ only; that is, statistically speaking we can give the proof under the conditional distribution of (X_1, X_2) given $(X_3, \cdots, X_n)$.

Assume that $a_1 > b_1 \geq b_2 > a_2$ and that $\mathbf{a} = (a_1, a_2) > \mathbf{b} = (b_1, b_2)$. For fixed $c = \sum_1^2 a_i = \sum_1^2 b_i$ let us consider the set

$$C(\mathbf{a}) = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, |x_i| \leq a_i, i = 1, 2\}.$$

Then the line $x_1 + x_2 = \lambda$ is tangent to the sets of boundary points of $C(\mathbf{a})$ and $C(\mathbf{b})$ for $\lambda = \pm c$. Hence Condition A1 is satisfied. Also, simple algebra shows that, for $\lambda \in [-c, c]$,

$$\| \ell_\lambda(\mathbf{a}) \| = \begin{cases} \sqrt{2}(c - |\lambda|) & \text{for } |\lambda| \geq a_1 - a_2, \\ \sqrt{8}a_2 & \text{otherwise,} \end{cases}$$

$$\delta_\lambda(\mathbf{a}) = \begin{cases} (a_1 - a_2)/\sqrt{2} & \text{for } |\lambda| \geq a_1 - a_2, \\ |\lambda|/\sqrt{2} & \text{otherwise.} \end{cases}$$

Therefore it can be verified that Condition A2 is also satisfied, and the proof follows from Lemma 2.1.

PROOF OF THEOREM 2.2 Again consider $n = 2$, and assume that $a_1^2 > b_1^2 \geq b_2^2 > a_2^2$, $c^2 = \sum_1^2 a_i^2 = \sum_1^2 b_i^2$ is kept fixed. Let

$$C(\mathbf{a}) = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^2, \textstyle\sum_1^2 (x_i/a_i)^2 \leq 1\}.$$

Then again the line $x_1 + x_2 = \lambda$ is tangent to the sets of boundary points of $C(\mathbf{a})$ and $C(\mathbf{b})$ for $\lambda = \pm c$. Moreover, it follows from elementary computations that

$$\| \ell_\lambda(\mathbf{a}) \| = \sqrt{8}\, a_1 a_2 (c^2 - \lambda^2)^{1/2}/c^2,$$

$$\delta_\lambda(\mathbf{a}) = (a_1^2 - a_2^2)\, |\lambda|/\sqrt{2}c^2.$$

Therefore Conditions A1 and A2 are also satisfied, and the theorem follows.

Combining Theorems 2.1 and 2.2 with a result in Marshall and Olkin (1979) we can obtain a seemingly more general statement. This is given below as corollaries.

COROLLARY 2.1. *If* $f(\mathbf{x})$, *the density of* $\mathbf{X}$, *is Schur-concave, and if* $\phi(a):[0,\infty)\to[0,\infty)$ *is strictly increasing and convex, then the rectangular probability* $P(|X_i|\le a_i, i=1,\cdots,n)$ *is a Schur-concave function of* $(\phi(a_1),\cdots,\phi(a_n))$. *That is,*

$$(\phi(a_1),\cdots,\phi(a_n))\succ(\phi(b_1),\cdots,\phi(b_n))$$

implies

$$P(|X_i|\le a_i, i=1,\cdots,n)\le P(|X_i|\le b_i, i=1,\cdots,n).$$

PROOF. Write the rectangular probability as

$$\gamma(\mathbf{a})=\gamma(\phi^{-1}(\phi(a_1)),\cdots,\phi^{-1}(\phi(a_n))).$$

Since $-\gamma$ is decreasing and (by Theorem 2.1) Schur-convex, and since ϕ^{-1} is increasing and concave, the result follows immediately from case (vi) of Table 2 in Marshall and Olkin (1979, page 63). □

COROLLARY 2.2. *If* $f(\mathbf{x})$ *and* $\phi(a)$ *satisfy the conditions given in Corollary* 2.1, *then the elliptical probability* $P\{\sum_1^n(X_i/a_i)^2\le 1\}$ *is a Schur-concave function of* $(\phi(a_1^2),\cdots,\phi(a_n^2))$.

PROOF. Similar to proof of Corollary 2.1. □

These corollaries can be applied to obtain inequalities when the sum $\sum_1^n\phi(a_i)$ or $\sum_1^n\phi(a_i^2)$ is held fixed. If ϕ is a linear function, then they reduce to the statements in Theorems 2.1 and 2.2. Corollary 2.1 also yields the following geometric inequality: The volume of the set $\{\mathbf{x}\,|\,\mathbf{x}\in\mathbb{R}^n, |x_i|\le a_i, i=1,\cdots,n\}$ is a Schur-concave function of $(\phi(a_1),\cdots,\phi(a_n))$, and is maximized (for fixed $\sum_1^n\phi(a_i)$) when $a_1=\cdots=a_n$. For $\phi(a)=a$ this result is, of course, well-known.

Generally speaking, Theorems 2.1 and 2.2 can be regarded as inequalities via the diversity of the scale parameters. That is, they say that if $f(\mathbf{x})$ is Schur-concave, then

$$\int_D f(a_1y_1,\cdots,a_ny_n)\prod_1^n d(a_iy_i) \tag{2.9}$$

is a Schur-concave function of $\mathbf{a}$ for

$$D=\{\mathbf{y}\,|\,\mathbf{y}\in\mathbb{R}^n, |y_i|\le 1, i=1,\cdots,n\} \tag{2.10}$$

and

$$D=\{\mathbf{y}\,|\,\mathbf{y}\in\mathbb{R}^n, \textstyle\sum_1^n y_i^2\le 1\}, \tag{2.11}$$

respectively; or equivalently, the function defined by

$$\psi(\mathbf{a})=Eg(X_1/a_1,\cdots,X_n/a_n) \tag{2.12}$$

is Schur-concave where g is the indicator function of D of (2.10) or of (2.11). This fact is analogous to, but different from, the inequality due to Marshall and Proschan (1965) which was obtained under the assumption of concavity instead of Schur-concavity. Note that, as pointed out in Tong (1980, page 116), the result of Marshall and Proschan (1965) does not yield probability inequalities.

A legitimate question is whether or not the results can be generalized to the class of convex sets

$$D_k=\{\mathbf{y}\,|\,\mathbf{y}\in\mathbb{R}^n, \textstyle\sum_1^n y_i^k\le 1\},\qquad k=2,4,6,\cdots,\infty, \tag{2.13}$$

for fixed

$$c^{k/(k-1)}=\textstyle\sum_1^n a_i^{k/(k-1)}=\sum_1^n b_i^{k/(k-1)}. \tag{2.14}$$

In view of the facts that the sets given in (2.10) and (2.11) are D_∞ (plus the boundary points) and D_2, respectively, and that for general k Condition A1 is already satisfied, it is conjectured that such an extension is possible. It also appears that the proof of Lemma 2.1 can be modified to obtain similar results for general symmetric and convex sets, but it is not yet known to the author to what extent such a generalization is possible.

3. Applications. In this section, we give some applications of the main results.

APPLICATION 3.1. *Independent and identically distributed random variables.* Let $X_1, \cdots, X_n$ be i.i.d. univariate random variables with density $h(x)$. If $h(x)$ is a log-concave function of x, then $\mathbf{X} = (X_1, \cdots, X_n)$ is Schur-concave, and Theorems 2.1 and 2.2 can be applied to obtain bounds on their joint probabilities.

APPLICATION 3.2. *Symmetric unimodal densities.* If $f(\mathbf{x})$ is permutation symmetric and unimodal, then it is Schur-concave, hence Theorems 2.1 and 2.2 apply.

APPLICATION 3.3 *Exchangeable multivariate normal and t variables.* If $\mathbf{X} = (X_1, \cdots, X_n)$ is a normal variable with means μ, variances σ^2 and correlations $\rho \in (-1/(n-1), 1)$, then its density is Schur-concave, and Theorems 2.1 and 2.2 apply. Moreover, if S is a $\sqrt{\chi^2_{(\nu)}/\nu}$ variable and is independent of $\mathbf{X}$, and if $\mu = 0$, $\sigma^2 = 1$, then the density of the multivariate t variables $\mathbf{t} = (X_1/S, \cdots, X_n/S)$ is also Schur-concave. Application 3.3 includes the following special cases.

Normal and t probability bounds. In many applied problems one needs to evaluate the value of a rectangular probability for exchangeable normal or t variables. Suppose that we find the table value from existing tables with all coordinates equal to $\bar{a} = \frac{1}{n}\sum_1^n a_i$, (e.g., using the tables in Dunn, Kronmal and Yee (1968)). Then from Theorem 2.1 we have

$$P(|X_i| \le a_i, i = 1, \cdots, n) \le P(|X_i| \le \bar{a}, i = 1, \cdots, n), \tag{3.1}$$

$$P(|t_i| \le a_i, i = 1, \cdots, n) \le P(|t_i| \le \bar{a}, i = 1, \cdots, n), \tag{3.2}$$

Thus as a special consequence the existing tables can be used to obtain numerical values of the upper bounds. To illustrate this point, consider exchangeable normal variables X_1, X_2, X_3, X_4 with $\mu = 0$, $\sigma^2 = 1$ and $\rho = ½$, and consider the possible configurations of the $\mathbf{a}$ vector:

$$(3.0, 3.0, 1.0, 1.0) \succ (2.75, 2.25, 1.75, 1.25) \succ (2.5, 2.5, 1.5, 1.5) \succ (2.0, 2.0, 2.0, 2.0).$$

Based on numerical calculations on an IBM 365/370 computer, the corresponding values of $\gamma(\mathbf{a})$ are

$$0.4975 < 0.7318 < 0.7613 < 0.8569.$$

Note that the value of $\gamma(2.0, 2.0, 2.0, 2.0)$ is an upper bound on $\gamma(\mathbf{a})$ for all $\mathbf{a}$ satisfying $\sum_{i=1}^4 a_i = 8.0$.

Distributions of sum squares of normal variables and linear combinations of central and noncentral Chi squared variables. Let $(Y_1, \cdots, Y_n)$ be a normal variable with variances σ_i^2, means $\mu\sigma_i (i = 1, \cdots, n)$ and correlations ρ. Then for all y, the probability $P(\sum_1^n Y_i^2 \le y)$ is a Schur-concave function of $(\phi(\sigma_1^{-2}), \cdots, \phi(\sigma_n^{-2}))$. Thus the maximum of the probability (for fixed $\sum \sigma_i^{-2}$) is attained when each of σ_i^2 is replaced by their harmonic mean σ_0^2. For the case $\mu = 0$ and $\rho = 0$, the upper bound on the probability is $P(\chi^2_{(n)} \le y/\sigma_0^2)$; this yields a bound for the linear combinations of independent Chi squared variables, and should be useful in the power consideration of certain tests. For the case $\mu \ne 0$ and $\rho = 0$, this yields a new inequality for the noncentral Chi squared distribution via the diversity of the variances.

APPLICATION 3.4. *Optimal allocation of observations.* Let there be n equally correlated (or independent) normal populations with means μ_i and known variances σ_i^2, and let $\bar{X}_i$ denote the sample mean with sample size N_i from the ith population ($i = 1, \cdots, n$). Let us denote (for $b_i > 0$, $i = 1, \cdots, n$)

$$E = \times_{i=1}^{n} [\bar{X}_i - b_i, \bar{X}_i + b_i]$$

to be the two-sided confidence region for $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)$. Then the confidence probability is $P(|X_i| \leq \sqrt{N_i}\, b_i/\sigma_i, i = 1, \cdots, n)$, where $(X_1, \cdots, X_n)$ is a normal variable with means 0, variances 1 and correlations ρ. A practical question of concern is this: Under the condition that $b_1/\sigma_1 = \cdots = b_n/\sigma_n$, what is the best allocation of observations so that, for fixed total sample size $N = \sum_{i=1}^{n} N_i$ (but not for fixed $\sum_{i=1}^{n} \sqrt{N_i}$), the confidence probability is maximized? This question can be answered immediately by choosing $\phi(a) = a^2$ in Corollary 2.1. Therefore, the best allocation is such that $|N_i - N_j| \leq 1$ for all $i \neq j$. Further, it yields a chain of inequalities, and says that the more diverse are $N_1, \cdots, N_n$ in the sense of majorization, the smaller is the coverage probability.

Acknowledgment. I wish to thank a referee who pointed out the present simple proof of Corollary 2.1.

REFERENCES

DUNN, O. J., KRONMAL, R. A., and YEE, W. J. (1968). Tables of the multivariate t distribution. Tech. Report, School of Public Health, UCLA, Los Angeles.

MARSHALL, A. W. and OLKIN, I. (1974). Majorization in multivariate distributions. *Ann. Statist.* **2** 1189–1200.

MARSHALL, A. W. and OLKIN, I., (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

MARSHALL, A. W. and PROSCHAN, F. (1965). An inequality for convex functions involving majorization. *J. Math. Anal. Appl.* **12** 87–90.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions.* Academic, New York.

DEPARTMENT OF MATHEMATICS AND STATISTICS
THE UNIVERSITY OF NEBRASKA-LINCOLN
LINCOLN, NEBRASKA 68588